

PCT

WELTORGANISATION FÜR GEISTIGES EIGENTUM

Internationales Büro

INTERNATIONALE ANMELDUNG VERÖFFENTLICH NACH DEM VERTRAG ÜBER DIE  
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)



(51) Internationale Patentklassifikation <sup>6</sup> : C12N 15/54, 9/12, 15/11, 15/85, C12Q 1/68, A01K 67/027		A2	(11) Internationale Veröffentlichungsnummer: WO 99/33998 (43) Internationales Veröffentlichungsdatum: 8. Juli 1999 (08.07.99)
(21) Internationales Aktenzeichen: PCT/EP98/08216 (22) Internationales Anmeldedatum: 22. Dezember 1998 (22.12.98)		(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(30) Prioritätsdaten: 197 57 984.1 24. Dezember 1997 (24.12.97) DE		Veröffentlicht Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.	
(71) Anmelder (für alle Bestimmungsstaaten ausser US): BAYER AKTIENGESELLSCHAFT [DE/DE]; D-51368 Leverkusen (DE).			
(72) Erfinder; und (75) Erfinder/Anmelder (nur für US): HAGEN, Gustav [DE/DE]; Bertha-von-Suttner-Strasse 31, D-51373 Leverkusen (DE). WICK, Maresa [DE/DE]; Andreas-Gryphius-Strasse 26, D-51065 Köln (DE). ZUBOV, Dmitry [RU/DE]; Roggendorfstrasse 59, D-51061 Köln (DE).			
(74) Gemeinsamer Vertreter: BAYER AKTIENGESELLSCHAFT; D-51368 Leverkusen (DE).			
(54) Title: REGULATORY DNA SEQUENCES OF THE HUMAN CATALYTIC TELOMERASE SUB-UNIT GENE, DIAGNOSTIC AND THERAPEUTIC USE THEREOF			
(54) Bezeichnung: REGULATORISCHE DNA-SEQUENZEN DES GENS DER HUMANEN KATALYTISCHEN TELOMERASE-UNTEREINHEIT UND DEREN DIAGNOSTISCHE UND THERAPEUTISCHE VERWENDUNG			
(57) Abstract			
The present invention relates to regulatory DNA sequences containing promotor sequences, in addition to intervening sequences, for the human catalytic telomerase sub-unit gene. The invention also relates to the use of said DNA sequences for pharmaceutical, diagnostic and therapeutic purposes, especially in the treatment of cancer and ageing.			

**(57) Zusammenfassung**

Diese Erfindung betrifft regulatorische DNA-Sequenzen, beinhaltend Promotorsequenzen, sowie Intronsequenzen, für das Gen der humanen kalytischen Telomerase-Untereinheit. Darüber hinaus betrifft diese Erfindung die Verwendung dieser DNA-Sequenzen für pharmazeutische, diagnostische und therapeutische Zwecke, vor allem in der Behandlung von Krebs und Alterung.

**LEDIGLICH ZUR INFORMATION**

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Lettland	SZ	Swasiland
AZ	Aserbaidschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mall	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		

Regulatorische DNA-Sequenzen des Gens der humanen katalytischen Telomerase-Untereinheit und deren diagnostische und therapeutische Verwendung

5 Aufbau und Funktion der Chromosomenenden

Das genetische Material eukaryontischer Zellen ist auf linearen Chromosomen verteilt. Die Enden der Erbanlagen werden, abgeleitet von den griechischen Wörtern *telos* (Ende) und *meros* (Teil, Segment), als Telomere bezeichnet. Die meisten 10 Telomere bestehen aus Wiederholungen von kurzen Sequenzen, die überwiegend aus Thymin und Guanin aufgebaut sind (Zakian, 1995). In allen bislang untersuchten Wirbeltieren werden die Telomere aus der Sequenz TTAGGG aufgebaut (Meyne *et al.*, 1989).

15 Die Telomere üben verschiedene wichtige Funktionen aus. Sie verhindern die Fusion von Chromosomen (McClintock, 1941) und damit die Entstehung von dizentrischen Erbanlagen. Solche Chromosomen mit zwei Centromeren können durch Verlust der Heterozygotie bzw. Verdopplung oder Verlust von Genen zur Entwicklung von Krebs führen.

20 Des Weiteren dienen Telomere dazu, intakte Erbanlagen von beschädigten zu unterscheiden. So stellten Hefezellen ihre Zellteilung ein, wenn sie ein Chromosom ohne Telomer enthielten (Sandell und Zakian, 1993).

25 Eine weitere wichtige Aufgabe erfüllen Telomere bei der DNA-Replikation eukaryontischer Zellen. Im Gegensatz zu den zirkulären Genomen von Prokaryonten können die linearen Chromosomen der Eukaryonten von dem DNA Polymerase-Komplex nicht vollständig repliziert werden. Zur Initiation der DNA-Replikation sind RNA-Primer notwendig. Nach Abspaltung der RNA-Primer, Verlängerung der 30 Okazaki-Fragmente und anschließender Ligation fehlt dem neu-synthetisierten DNA-Strang das 5'-Ende, denn dort kann der RNA-Primer nicht durch DNA ersetzt

werden. Ohne besondere Schutzmechanismen würden daher die Chromosomen mit jeder Zellteilung schrumpfen ("end-replication problem"; Harley *et al.*, 1990). Die nicht-kodierenden Telomersequenzen stellen vermutlich eine Pufferzone dar, um dem Verlust von Genen vorzubeugen (Sandell und Zakian, 1993).

5

Darüberhinaus spielen Telomere auch eine wichtige Rolle bei der Regulation der zellulären Alterung (Olovnikov, 1973). Humane somatische Zellen zeigen in Kultur eine limitierte Replikationskapazität; sie werden nach einer gewissen Zeit seneszent. In diesem Zustand teilen sich die Zellen selbst nach Stimulierung mit Wachstumsfaktoren nicht mehr, sterben aber nicht, sondern bleiben metabolisch aktiv (Goldstein, 1990). Verschiedene Beobachtungen sprechen für die Hypothese, daß eine Zelle anhand der Länge ihrer Telomere bestimmt, wie oft sie sich noch teilen kann (Allsopp *et al.*, 1992).

10

Zusammenfassend besitzen die Telomere somit zentrale Funktionen bei der Alterung von Zellen sowie der Stabilisierung des genetischen Materials und Verhinderung von Krebs.

#### Das Enzym Telomerase synthetisiert die Telomere

15

Wie oben beschrieben können Organismen mit linearen Chromosomen ohne einen speziellen Schutzmechanismus ihr Genom nur unvollständig replizieren. Die meisten Eukaryonten verwenden zur Regeneration der Telomersequenzen ein spezielles Enzym, die Telomerase. In den bislang untersuchten Einzellern wird Telomerase konstitutiv exprimiert. Dagegen wurde in Menschen die Telomerase-Aktivität nur in Keimzellen und Tumorzellen gemessen, wogegen benachbartes somatisches Gewebe keine Telomerase enthielt (Kim *et al.*, 1994).

20

Funktionell kann die Telomerase auch als terminale Telomertransferase bezeichnet werden, die als Multiproteinkomplex im Zellkern lokalisiert ist. Während der RNA-Anteil der humanen Telomerase schon seit längerem bekannt ist (Feng *et al.*, 1995),

wurde kürzlich die katalytische Untereinheit dieser Enzymgruppe in verschiedenen Organismen identifiziert (Lingner *et al.*, 1997; vgl. unsere ebenfalls anhängige Anmeldung PCT EP/98/03468). Diese katalytischen Untereinheiten der Telomerase sind sowohl untereinander als auch zu bisher allen bekannten reversen Transkriptasen 5 auffällig homolog.

Auch in WO 98/14592 werden Nukleinsäure- und Aminosäuresequenzen der katalytischen Telomerase-Untereinheit beschrieben.

10 Aktivierung der Telomerase in menschlichen Tumoren

Eine Aktivität der Telomerase konnte in Menschen ursprünglich nur in Keimbahnzellen, nicht aber in normalen somatischen Zellen (Hastie *et al.*, 1990; Kim *et al.*, 1994) nachgewiesen werden. Nach der Entwicklung eines sensitiveren Nachweisverfahrens 15 (Kim *et al.*, 1994) wurde auch in hematopoietischen Zellen eine geringe Telomeraseaktivität detektiert (Broccoli *et al.*, 1995; Counter *et al.*, 1995; Hiyama *et al.*, 1995). Allerdings wiesen diese Zellen trotzdem eine Reduktion der Telomere auf (Vaziri *et* 20 *al.*, 1994; Counter *et al.*, 1995). Noch ist nicht geklärt, ob die Menge an Enzym in diesen Zellen nicht ausreichend für eine Kompensation des Telomerverlustes ist, oder ob die gemessene Telomerase-Aktivität von einer Subpopulation, z.B. unvollständig ausdifferenzierten CD34<sup>+</sup>38<sup>+</sup>-Vorläuferzellen, herrührt (Hiyama *et al.*, 1995). Zur Klärung wäre ein Nachweis der Telomerase-Aktivität in einer einzelnen Zelle nötig.

Interessanterweise wurde jedoch in einer großen Zahl der bislang getesteten Tumorgewebe 25 eine signifikante Telomerase-Aktivität nachgewiesen (1734/2031, 85 %; Shay, 1997), während in normalem somatischen Gewebe keine Aktivität gefunden wurde (1/196, <1 %, Shay, 1997). Verschiedene Untersuchungen zeigten außerdem, daß in seneszenten Zellen, die mit viralen Oncoproteinen transformiert wurden, die Telomere weiterhin schrumpften und Telomerase nur in der Subpopulation entdeckt 30 werden konnte, die die Wachstumskrise überlebte (Counter *et al.*, 1992). In diesen immortalisierten Zellen waren auch die Telomere stabil (Counter *et al.*, 1992). Ähnli-

che Befunde aus Untersuchungen an Mäusen (Blasco *et al.*, 1996) stützen die Annahme, daß eine Reaktivierung der Telomerase ein spätes Ereignis in der Tumorgeneze ist.

5 Basierend auf diesen Ergebnissen wurde eine "Telomerase-Hypothese" entwickelt, die den Verlust von Telomersequenzen und Zellalterung mit der Aktivität von Telomerase und der Entstehung von Krebs verbindet. In langlebigen Spezies wie dem Menschen kann das Schrumpfen der Telomere als ein Mechanismus zur Tumorsuppression angesehen werden. Ausdifferenzierte Zellen, die keine Telomerase enthalten, stellen bei einer bestimmten Länge der Telomere ihre Zellteilung ein. Mutiert eine solche Zelle, so kann aus ihr nur dann ein Tumor entstehen, wenn die Zelle ihre Telomere verlängern kann. Ansonsten würde die Zelle weiterhin Telomersequenzen verlieren, bis ihre Chromosomen instabil werden und sie schließlich zugrunde geht. Die Reaktivierung der Telomerase ist vermutlich der Hauptmechanismus von Tumorzellen zur Stabilisation ihrer Telomere.

20 Aus diesen Beobachtungen und Überlegungen ergibt sich, daß eine Inhibition der Telomerase eine Therapie von Tumoren erlauben sollte. Konventionelle Krebstherapien mit Zytostatika oder kurzweligen Strahlen schädigen nicht nur die Tumorzellen, sondern alle sich teilenden Zellen des Körpers. Da aber außer Tumorzellen nur Keimbahnzellen eine signifikante Telomerase-Aktivität enthalten, würden Telomerase-Inhibitoren spezifischer die Tumorzellen angreifen und somit weniger unerwünschte Nebenwirkungen hervorrufen. In allen bislang getesteten Tumorgeweben wurde eine Telomerase-Aktivität nachgewiesen, so daß diese Therapeutika gegen alle Krebsarten eingesetzt werden könnten. Die Wirkung von Telomerase-Inhibitoren würde dann eintreten, wenn die Telomere der Zellen sich soweit verkürzt haben, daß das Genom instabil wird. Da Tumorzellen meist kürzere Telomere aufweisen als normale somatische Zellen, würden zuerst Krebszellen durch Telomerase-Inhibitoren eliminiert werden. Zellen mit langen Telomeren, wie die Keimzellen, würden dagegen erst viel später geschädigt werden. Telomerase-Inhibitoren stellen somit einen zukunftsweisenden Weg für die Therapierung von Krebs dar.

Eindeutige Antworten auf die Frage nach der Art und den Angriffspunkten physiologischer Telomerase-Inhibitoren werden möglich sein, wenn auch die Regulation der Genexpression der Telomerase identifiziert ist.

5

#### Regulation der Genexpression in Eukaryonten

Die eukaryotische Genexpression, d.h. der zelluläre Informationsfluß von der DNA über die RNA zum Protein, weist vielfältige Ansatzpunkte für regulatorische Mechanismen auf. Einzelne Kontrollstufen sind z.B. die Gen-Amplifikation, Rekombination von Genloci, Chromatinstruktur, DNA-Methylierung, Transkription, posttranskriptionelle mRNA-Modifikationen, mRNA-Transport, Translation und post-transkriptionale Proteinmodifikationen. Nach bisherigen Studien besitzt die Kontrolle auf der Ebene der Transkriptionsinitiation die größte Bedeutung (Latchman, 1991).

15

Unmittelbar stromaufwärts vom Transkriptionsstart eines von der RNA-Polymerase II transkribierten Gens liegt eine Region, die für die Steuerung der Transkription verantwortlich ist und als Promotorregion bezeichnet wird. Ein Vergleich der Nukleotidsequenzen von Promotorregionen vieler bekannter Gene zeigt, daß bestimmte Sequenzmotive in dieser Region häufig vorkommen. Zu diesen Elementen gehören unter anderem die TATA-Box, die CCAAT-Box und die GC-Box, die von spezifischen Proteinen erkannt werden. Die TATA-Box, die etwa 30 Nukleotide stromaufwärts vom Transkriptionsstart entfernt positioniert ist, wird z.B. von der TFIID-Untereinheit TBP („TATA-box binding protein“) erkannt, wogegen bestimmte GC-reiche Sequenzabschnitte vom Transkriptionsfaktor Sp1 („specificity protein1“) spezifisch gebunden werden.

20

Funktionell kann man den Promotor in einen regulativen und einen konstitutiven Abschnitt unterteilen (Latchman, 1991). Der konstitutive Kontrollbereich umfaßt den sogenannten Kernpromotor („corepromoter“), der die korrekte Initiation der Transkription ermöglicht. Er enthält die als UPE's (upstream promoter elements) be-

25

30

schriebenen Sequenzelemente, die für eine effiziente Transkription notwendig sind. Die regulativen Kontrollabschnitte, die mit den UPE's verflochten sein können, weisen Sequenzelemente auf, die an der signalabhängigen Regulation der Transkription durch Hormone, Wachstumsfaktoren usw. beteiligt sein können. Sie ver-  
mitteln gewebs- oder zellspezifische Promotoreigenschaften.

Ein charakteristisches Merkmal eukaryotischer Gene sind DNA-Abschnitte, die über vergleichsweise große Distanzen hinweg Einfluß auf die Genexpression nehmen können. Diese Elemente können stromaufwärts, stromabwärts oder innerhalb einer  
10 Transkriptionseinheit lokalisiert sein und unabhängig von ihrer Orientierung ihre Funktion wahrnehmen. Diese Sequenzabschnitte können die Promotoraktivität verstärken (Enhancer) oder abschwächen (Silencer). Ähnlich wie die Promotorregionen beherbergen auch Enhancer und Silencer mehrere Bindungsstellen für Transkriptionsfaktoren.

15 Die Erfindung betrifft die DNA-Sequenzen aus der 5'-flankierenden Region des Gens der katalytisch aktiven humanen Telomerase-Untereinheit sowie Intron-Sequenzen für dieses Gen.

20 Die Erfindung betrifft insbesondere die 5'-flankierende regulatorische DNA-Sequenz, enthaltend die Promotor-DNA-Sequenz für das Gen der humanen katalytischen Telomerase Untereinheit gemäß Fig. 10 (SEQ ID NO 3).

25 Die Erfindung betrifft weiterhin regulatorisch wirksame Teilbereiche der 5'-flankierenden regulatorischen DNA-Sequenz gemäß Fig. 4 (SEQ ID NO 1).

Weiterhin sind Gegenstand der vorliegenden Erfindung Intron-Sequenzen für das  
30 Gen der humanen katalytischen Telomerase-Untereinheit, insbesondere solche, die regulatorische Wirkung haben. Die erfindungsgemäßen Intronsequenzen werden im Rahmen von Beispiel 5 detailliert beschrieben (vgl. SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und 20).

Die Erfindung betrifft weiterhin ein rekombinantes Konstrukt, das die erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierende DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereiche davon  
5 beinhaltet.

Bevorzugt sind rekombinante Konstrukte, die neben den erfindungsgemäßen DNA-Sequenzen, insbesondere der 5'-flankierenden DNA-Sequenz des Gens der humanen katalytischen Telomerase Untereinheit oder Teilbereichen davon, eine oder mehrere  
10 weitere DNA-Sequenzen, die für Polypeptide oder Proteine kodieren, enthalten.

Gemäß einer besonders bevorzugten Ausführungsform kodieren diese weiteren DNA-Sequenzen für antitumorale Proteine.

15 Besonders bevorzugte antitumorale Proteine sind solche, die die Angiogenese direkt oder indirekt inhibieren. Zu diesen Proteinen zählen beispielsweise:

Plasminogenaktivatorinhibitor (PAI-1), PAI-2, PAI-3, Angiostatin, Endostatin, Platelet factor 4, TIMP-1, TIMP-2, TIMP-3, Leukemia Inhibitory Factor (LIF).

20 Ebenfalls besonders bevorzugt sind antitumorale Proteine, welche direkt oder indirekt eine zytostatische Wirkung auf Tumoren aufweisen. Hierzu zählen im besonderen:

25 Perforin, Granzym, IL-2, IL-4, IL-12, Interferone, wie beispielsweise IFN- $\alpha$ , IFN- $\beta$ , IFN- $\gamma$ , TNF, TNF- $\alpha$ , TNF- $\beta$ , Oncostatin M; Tumorsuppressorgene, wie z.B. p53, Retinoblastoma.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche gegebenenfalls  
30 zusätzlich zur antitumoralen Wirkung Entzündungen stimulieren und hierdurch zur Elimination von Tumorzellen beitragen. Hierzu zählen beispielsweise:

RANTES, Monocyte chemotactic and activating factor (MCAF), IL-8, Macrophage inflammatory protein (MIP-1 $\alpha$ , $\beta$ ), Neutrophil activating protein-2 (NAP-2), IL-3, IL-5, human leukemia inhibitory factor (LIF), IL-7, IL-11, IL-13, GM-CSF, G-CSF,

5 M-CSF.

Weiterhin besonders bevorzugt sind antitumorale Proteine, welche aufgrund ihrer Wirkung als Enzyme in der Lage sind, Vorstufen eines antitumoralen Wirkstoffes in einen antitumoralen Wirkstoff zu überführen. Zu diesen Enzymen zählen beispiels-

10 weise:

Herpes Simplex Virus Thymidinkinase, Varizella Zoster Virus Thymidinkinase, bakterielle Nitroreductase, bakterielle  $\beta$ -Glukuronidase, pflanzliche  $\beta$ -Glukuronidase aus *Secale careale*, humane Glukuronidase, humane Carboxypeptidase, bakterielle Carboxypeptidase, bakterielle  $\beta$ -Lactamase, bakterielle Cytosindeaminidase, humane Katalase bzw. Phosphatase, humane alkalische Phosphatase, Typ 5 saure Phosphatase, humane Lysooxidase, humane saure D-Aminooxidase, humane Glutathion Peroxidase, humane Eosinophilen Peroxidase, humane Schilddrüsen Peroxidase.

15 20 Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für Faktor VIII, IX oder Teilfragmente davon kodieren. Zu diesen DNA-Sequenzen zählen auch andere Blutgerinnungsfaktoren

25 Die obengenannten rekombinanten Konstrukte können auch DNA-Sequenzen enthalten, die für ein Reporterprotein kodieren. Zu diesen Reporterproteinen zählen beispielsweise:

30 Chloramphenicolacetyltransferase (CAT), Glühwürmchen Luziferase (LUC),  $\beta$ -Galaktosidase ( $\beta$ -Gal), Sezernierte alkalische Phosphatase (SEAP), Humanes Wachstumshormon (hGH),  $\beta$ -Glukuronidase (GUS), Grün-fluoreszierendes Protein (GFP) und alle davon abgeleiteten Varianten, Aquarin, Obelin.

Erfindungsgemäße rekombinante Konstrukte können auch DNA kodierend für die humane katalytische Telomerase Untereinheit und deren Varianten und Fragmente in antisense Orientierung enthalten. Gegebenenfalls können diese Konstrukte auch andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente in antisense Orientierung enthalten.

Die rekombinanten Konstrukte können neben der DNA, kodierend für die humane katalytische Telomerase Untereinheit, sowie deren Varianten und Fragmente auch andere Protein-Untereinheiten der humanen Telomerase und die Telomerase-RNA-Komponente enthalten.

15 Die Erfindung betrifft weiterhin einen Vektor, enthaltend die oben genannten erfindungsgemäßen DNA-Sequenzen, insbesondere die 5'-flankierenden DNA-Sequenzen, sowie eine oder mehrere der oben genannten anderen DNA-Sequenzen.

Bevorzugter Vektor für solche Konstrukte ist ein Virus, beispielsweise ein Retrovirus, Adenovirus, adeno-assoziiertes Virus, Herpes Simplex Virus, Vaccina Virus, lentivirales Virus, Sindbis Virus und ein Semliki Forest Virus.

20 Ebenfalls bevorzugt sind Plasmide als Vektoren.

Die Erfindung betrifft weiterhin pharmazeutische Präparate, enthaltend erfindungsgemäße rekombinante Konstrukte bzw. Vektoren; beispielsweise eine Zubereitung in einem kolloidalen Dispersionssystem.

Geeignete kolloidale Dispersionssysteme sind beispielsweise Liposome oder Polylysin-Liganden.

30 Die Zubereitungen der erfindungsgemäßen Konstrukte bzw. Vektoren in kolloidalen Dispersionssystemen können um einen Liganden ergänzt sein, der an Membranstruk-

turen von Tumorzellen bindet. Ein solcher Ligand kann z.B. an das Konstrukt bzw. den Vektor angeknüpft sein oder auch Bestandteil der Liposomenstruktur sein.

Geeignete Liganden sind insbesondere polyklonale oder monoklonale Antikörper

5 oder Antikörperfragmente hiervon, die mit ihren variablen Domänen an Membranstrukturen von Tumorzellen binden, oder endständige Mannose-tragende Substanzen, Zytokine, Wachstumsfaktoren oder Fragmente bzw. Teilesequenzen hiervon, die an Rezeptoren auf Tumorzellen binden.

10 Entsprechende Membranstrukturen sind beispielsweise Rezeptoren für ein Zytokin oder einen Wachstumsfaktor, wie z.B. IL-1, EGF, PDGF, VEGF, TGF  $\beta$ , Insulin oder Insulin-like Growth Factor (ILGF), oder Adhäsionsmoleküle, wie z. B. SLeX, LFA-1, MAC-1, LECAM-1 oder VLA-4, oder der Mannose-6-Phosphat-Rezeptor.

15 Zur vorliegenden Erfindung gehören pharmazeutische Zubereitungen, die neben den erfindungsgemäßen Vektorkonstrukten auch nichttoxische, inerte, pharmazeutisch geeignete Trägerstoffe enthalten können. Vorstellbar sind die Applikation (z.B. intravenös, intraarteriell, intramuskulär, subkutan, intradermal, anal, vaginal, nasal, transdermal, intraperitoneal, als Aerosol oder oral) am Ort eines Tumors oder die systemische Applikation dieser Zubereitungen.

Die erfindungsgemäßen Vektorkonstrukte können in der Gentherapie eingesetzt werden.

25 Die Erfindung betrifft weiterhin eine rekombinante Wirtszelle, insbesondere eine rekombinante eukaryotische Wirtszelle, enthaltend die vorstehend beschriebenen Konstrukte bzw. Vektoren.

30 Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit beeinflussen, wobei dieses Verfahren folgende Schritte umfaßt:

A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend die erfindungsgemäße regulatorische DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon, funktionell verknüpft mit einem Reportergen,

5

B. Messung des Substanzeffektes auf die Reportergenexpression.

10 Das Verfahren kann eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit verstärken.

15 Das Verfahren kann weiterhin eingesetzt werden zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der katalytischen Telomerase Untereinheit inhibieren.

20 Die Erfindung betrifft weiterhin ein Verfahren zur Identifizierung von Faktoren, die spezifisch an Fragmente der erfindungsgemäßen DNA-Fragmente, insbesondere der 5'-flankierenden regulatorischen DNA-Sequenz der katalytischen Telomerase Untereinheit, binden. Diese Methode beinhaltet ein Screening einer Expressions-cDNA-Bibliothek mit der vorstehend beschriebenen DNA-Sequenz oder Teilfragmenten unterschiedlichster Länge als Sonde.

25 Die vorstehend beschriebenen Konstrukte bzw. Vektoren können auch zur Herstellung transgener Tiere verwendet werden.

Die Erfindung betrifft weiterhin ein Verfahren zur Detektion Telomerase-assozierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

30

A. Inkubation eines Konstruktes bzw. Vektors, enthaltend die erfindungsgemäße DNA-Sequenz, insbesondere die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit oder einen regulatorisch wirksamen Teilbereich davon sowie ein Reportergen mit Körperflüssigkeiten oder zellulären Proben,

5

B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten;

10 C. Vergleich des diagnostischen Werts mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe;

15 Detektion diagnostischer Werte, die höher oder niedriger als Standardvergleichswerte liegen, indiziert einen Telomerase-assozierten Zustand, der wiederum einen pathogenen Zustand indiziert.

Erläuterung der Abbildungen:

20 Fig. 1: Southern Blot-Analyse mit genomischer DNA verschiedener Spezies

A: Foto eines Ethidiumbromid gefärbten 0,7 %igen Agarosegels mit etwa 4 µg Eco RI geschnittener genomischer DNA. Die Spur 1 enthält Hind III geschnittene λ-DNA als Größenmarker (23,5, 9,4, 6,7, 4,4, 2,3, 2,0, und 0,6 kb). Die Spuren 2 bis 10 enthalten genomische DNA von Mensch, Rhesusaffe, Spraque Dawley Ratte, BALB/c Maus, Hund, Rind, Kaninchen, Huhn und Hefe (*Saccharomyces cerevisiae*).

25

B: Zu Fig.1 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse, hybridisiert mit einer radioaktiv-markierten etwa 720 bp langen hTC-cDNA Sonde.

30

Fig. 2: Restriktionsanalyse der rekombinanten  $\lambda$ -DNA des Phagenklons P12, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

5 Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,4 %igen Agarosegels. Die Spuren 1 und 2 enthalten Eco RI/Hind III geschnittene  $\lambda$ -DNA bzw eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 3 - 7 enthalten 250 ng mit Bam HI (Spur 3), Eco RI (Spur 4), Sal I (Spur 5), Xho I (Spur 6) und Sac I (Spur 7) geschnittene DNA des 10 rekombinanten Phagens. Die Pfeile kennzeichnen die zwei  $\lambda$ -Arme des Vektors EMBL3 Sp6/T7.

15 Fig. 3: Restriktionsanalyse und Southern Blot-Analyse der rekombinanten  $\lambda$ -DNA des Phagenklons, der mit einer Sonde aus dem 5'-Bereich der hTC-cDNA hybridisiert.

20 A: Die Abbildung zeigt ein Foto eines Ethidiumbromid gefärbten 0,8%igen Agarosegels. Die Spuren 1 und 15 enthalten eine 1 kb Leiter der Firma Gibco als Größenmarker. Die Spuren 2 bis 14 enthalten 250 ng geschnittene  $\lambda$ -DNA vom rekombinanten Phagenklon. Als Enzyme wurden eingesetzt: Spur 2: Sac I, Spur 3: Xho I, Spur 4: Xho I, Xba I, Spur 5: Sac I, Xho I, Spur 6: Sal I, Xho I, Xba I, Spur 7: Sac I, Xho I, Xba I, Spur 8: Sac I, Sal I, Xba I, Spur 9: Sac I, Sal I, BamH I, Spur 10: 25 Sac I, Sal I, Xho I, Spur 11: Not I, Spur 12: Sma I, Spur 13: leer, Spur 14: nicht verdaut.

30 B: Zu Fig.3 A korrespondierendes Autoradiogramm einer Southern Blot-Analyse. Als Sonde für die Hybridisierung wurde ein etwa 420 bp langes 5'-hTC-cDNA Fragment eingesetzt.

Fig. 4: Partielle DNA-Sequenz der 5'-flankierenden Region und des Promotors vom Gen der humanen katalytischen Telomerase-Untereinheit. Das ATG-Startcodon ist in der Sequenz fett hervorgehoben. Die dargestellte Sequenz entspricht SEQ ID NO 1.

5

Fig. 5: Identifizierung des Transkriptionsstarts durch Primer Extension-Analyse.

Die Abbildung zeigt ein Autoradiogramm eines denaturierenden Polyacrylamidgels, welches zur Darstellung einer Primer Extension-Analyse gewählt wurde. Als Primer wurde ein Oligonukleotid mit der Sequenz 10 5' GTTAAGTTGTTAGCTTACACTGGTTCTC 3' benutzt. In der Spur 1 wurde die Primer Extension Reaktion aufgetragen. Die Spuren G, A, T, C, stellen die Sequenzreaktionen mit dem gleichen Primer und den entsprechenden Dideoxynukleotiden dar. Der fette Pfeil kennzeichnet den 15 Haupt-Transkriptionsstart, die dünnen Pfeile weisen auf drei Neben-Transkriptionsstartpunkte hin.

Fig. 6: cDNA Sequenz der humanen katalytischen Telomerase-Untereinheit (hTC; vgl. unsere anhängige Anmeldung PCT/EP/98/03468). Die 20 dargestellte Sequenz entspricht SEQ ID NO 2.

Fig. 7: Strukturelle Organisation und Restriktionsmappe des humanen hTC-Gens und dessen 5'- und 3'-flankierende Region.

25 Exons sind als durchnummierete schwarz ausgefüllte Rechtecke und Introns als nicht ausgefüllte Bereiche hervorgehoben. Nichttranslatierte Sequenzabschnitte in den Exons sind schraffiert. Die Translation startet in Exon 1 und endet in Exon 16. Restriktionsenzymschnittstellen sind 30 wie folgt gekennzeichnet: S, SacI; X, Xhol. Die relative Anordnung der fünf Phagenklone (P2, P3, P5, P12, P17) und des Produktes aus dem „Genomic walking“ sind durch dünne Linien hervorgehoben. Wie durch

die Punkte gekennzeichnet, ist die Sequenz von Intron 16 nur teilweise entschlüsselt.

Fig. 8: HTL Splicevarianten.

5

A: Schematische Struktur der hTC mRNA Splicevarianten. Die vollständige hTC mRNA ist als grau unterlegtes Rechteck im oberen Bereich der Abb. dargestellt. Die 16 Exons sind entsprechend ihrer Größe dargestellt. Der Translationsstart (ATG) und das Stop-Codon, sowie das Telomerase-spezifische T-Motiv und die sieben RT-Motive sind hervorgehoben. Die hTC-Varianten sind in Deletions- und Insertionsvarianten unterteilt. In den Deletionen sind die fehlenden Exonsequenzen markiert. Die Insertionen sind durch zusätzliche weiße Rechtecke hervorgehoben. Größe und Herkunft der insertierten Sequenzen sind angegeben. Neu entstandene Stop-Codons sind markiert. Die Größe der Insertion von Variante INS2 ist unbekannt.

10

B: Exon Intron Übergänge der hTC-Splice-Varianten. Nichtgesplicte 5'- und 3'-flankierte Sequenzen sind als weiße Rechtecke hervorgehoben. Die Herkunft der Exon und Intron Sequenzen ist angegeben. Intron und Exon Sequenzen sind in Kleinbuchstaben, bzw. in Großbuchstaben dargestellt. Die Donor und Akzeptor Sequenzen der Splicestellen sind als graue Rechtecke unterlegt und deren Exon Intron Herkunft ist ebenfalls angegeben.

15

Fig. 9: Identifizierung des Transkriptionsstarts durch RT-PCR Analyse.

Die RT-PCR wurde mit cDNA-Bibliothek aus HL 60 Zellen und genomischer DNA als Positivkontrolle durchgeführt. Ein gemeinsamer 3'-Primer hybridisiert an eine Sequenzregion aus Exon 1. Die Position der verschiedenen 5' Primer in der kodierenden Region oder der 5'-flankierenden Region ist angegeben. In der Negativkontrolle wurde keine

20

25

Template-DNA in der PCR-Reaktion zugegeben. M: DNA-Größenmarker.

Fig. 10: Nukleotidsequenz und Strukturmerkmale des hTC-Promotors.

5 11273 bp der 5'-flankierenden hTC Gensequenz, beginnend mit dem Translationsstartcodon ATG (+1) sind dargestellt. Die putative Region des Translationsstarts ist unterstrichen. Mögliche regulatorische Sequenzabschnitte innerhalb der 4000 bp stromaufwärts des Translationsstarts sind umrandet. Die dargestellte Sequenz entspricht SEQ ID NO 3.

10

Fig. 11: Aktivität des hTC-Promotors in HEK-293 Zellen.

15 Im oberen Bereich der Abbildung sind die ersten 5000 bp der 5'-flankierenden hTC Genregion schematisch dargestellt. Das ATG-Startcodon ist hervorgehoben. CpG reiche Inseln sind durch graue Rechtecke markiert. Auf der linken Abbildungsseite sind die Größen der hTC Promotor-Luziferase Konstrukte dargestellt. Das pomerlose pGL2-Basic Konstrukt und das SV40 Promotorkonstrukt pGL2-Pro wurden in jeder Transfektion als Kontrollen eingesetzt. Auf der rechten Abbildungsseite sind ist die relative Luziferaseaktivität der verschiedenen Promotorkonstrukte in HEK-Zellen als durchgehende Balken gezeigt. Die Standardabweichung ist angegeben. Die Zahlenwerte repräsentieren den Durchschnitt von zwei unabhängigen Experimenten, die in Duplikaten durchgeführt wurden.

20

25

Tab. 1: Exon Intron Übergänge des hTC-Gens

Aufgelistet sind die Nukleotidsequenzen an den 3'- und 5' Spliceübergängen des hTC-Gens. Die Konsensussequenzen für Donor und Akzeptorsequenzen (AG und GT) sind durch graue Rechtecke unterlegt. Intronsequenzen (Kleinbuchstaben) und Exonsequenzen (Großbuchstaben), die die Spliceakzeptor- und Donorstellen flankieren sind gezeigt. Die Größe der Exons und Introns ist in bp angegeben.

30

Tab. 2: Potentielle Bindungsstellen für DNA-bindende Faktoren in der Nukleotidsequenz von Intron 2

5 Die Suche nach möglichen DNA-bindenden Faktoren (z.B. Transkriptionsfaktoren) wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Aufgelistet sind die Abkürzungen der identifizierten DNA-bindenden Faktoren und deren Lokalisation in  
10 Intron 2.

Tab. 1

3' Acceptor Sequence		5' Donor Sequence					
Intron	Exon	Exon	bp	Exon	Intron	Intron	bp
		No.		No.	No.	No.	
<b>5' Flankierende Region</b>							
caggggccttcccccggtag		1	281	CGCCCCCTCTTCGCCAG	gtggggcctcccccgggtcg	1	104
catgtcccttcgttttag		2	1354	TGGCTGCGCAGGCCAG	gttggggagggtgtggccgt	2	8616
gagggggctctatggcag		3	196	TGCAAAGCAATGGAATCG	gtactgttatccccacggca	3	2089
cccatatgtgtttcccgccag		4	181	GTTCCGGAGAGAAAGAGG	gtgggtgtgtttttttttta	4	687
ctccgccttcactacacag		5	180	TGAGCTGTACTTTGTCAAG	gtggggtgccggggacccccc	5	494
ctggatgtgtacggggcgt		6	156	CAAGGCCTTCAAAGCCAC	gtaaagggttcaacgtgtata	6	>4660
gtctctacccatgtggacc		7	96	TGGCGTGTATCGGCA	gtcttggccactgtccctgcga	7	980
AGCTCCCTCTGATTAAG		8	86	CCGTGGCATCGGGGCAA	gttgaggatcagggtggccagg	8	2481
GTCTTAAGTCCAGTGGCAG		9	114	CGGGGATTGGGGGAGCGG	gttgaggccctcttotcccc	9	1984
GCTGCTCTGCGTTCGGTG		10	72	ACGGGAAACCTTCCTCAG	gtggaggccccggccgtgtg	10	1871
GACCCCTGGTCCGAGGTGTC		11	189	TGGAGGGACTACTCCAG	gttgaggccacccgtggccgga	11	3801
CTATGCCGGACCTCCATC		12	127	CTGTGTTCTGGATTTCAG	gttgaggcaggctgtatggta	12	880
GTGAAAGCCATGAGACGG		13	62	TCCCTGTCAGGGGTACAG	gtgaggcccccaccaagggg	13	3187
GTTTCTGGCATGGTGTG		14	125	CTGAAAGCCAAAGACGGAG	gtatgtgcagggtgcctggc	14	781
GGATGCGCTGGGGCAA		15	138	CTGGGGTCACTCAGGACAG	gcaagggtgtgggtggaggcc	15	536
CCAGAGCGAGCTGAGTGC		16	664	TTTTTCAGTTTGTAAAAA	3' flankierende Region		
tctgatttttggcccgtag							

Tab. 2

Faktoren	Lokalisation in Intron 2
C/EBP	2925
CRE.2	2749
Spl	2378, 4094, 4526, 4787, 4835, 4995
AP-2 CS3	5099
AP-2 CS4	2213, 3699, 4667, 5878, 5938, 6059, 6180, 6496
AP-2 CS5	5350, 5798, 5880, 5940, 6061, 6182, 6375, 6498
PEA3	934, 2505
P53	2125
GR uteroglobin	848, 1487, 2956
PR uteroglobin	3331
Zeste-white	1577, 1619, 1703, 1745, 1787, 1829, 1871, 1913, 1955, 1997, 2039, 2081, 3518, 3709, 4765, 5014, 5055
GRE	846
MyoD-MCK right site/rev	447, 509, 558, 1370, 1595, 1900, 2028, 2099, 4557
MyoD-MCK left site	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
Ets-1 CS	6408
API	3784, 4406
CREB	2801
GATA-1	839, 1390, 3154
c-Myc	108, 118, 453, 1566, 1608, 1692, 1734, 1818, 1902, 1986, 2372, 2460, 2720, 3491, 5030
CACCC site	991
CCAAT site	1224
CCAC box	992
CAAT site	463, 2395
Rb site	992, 4663
TATA	3650
CDEI	106, 1564, 1606, 1690, 1732, 1816, 1900, 1984

Beispiele

Das menschliche Gen für die katalytische Telomerase Untereinheit (ghTC), sowie die 5' und 3' liegenden Bereiche dieses Gens wurden kloniert, der Startpunkt der 5 Transkription bestimmt, potentielle Bindungsstellen für DNA-bindende Proteine identifiziert, sowie aktive Promotorfragmente aufgezeigt. Die Sequenz der hTC-cDNA (Fig. 6) ist bereits in unserer ebenfalls anhängigen Anmeldung PCT/EP/98/03468 beschrieben. Wenn nicht gesondert erwähnt, beziehen sich sämtliche Angaben zur cDNA-Position auf diese Sequenz.

10

Beispiel 1

Durch eine genomische Southern Blot-Analyse wurde bestimmt, ob ghTC im menschlichen Genom ein Einzelgen darstellt oder mehrere Loci für das hTC-Gen 15 bzw. eventuell auch ghTC-Pseudogene existieren.

Hierzu wurde ein kommerziell erhältlicher Zoo-Blot der Firma Clontech einer Southern Blot-Analyse unterzogen. Dieser Blot enthält 4 µg Eco RI geschnittene 20 genomische DNA von neun verschiedenen Spezies (Mensch, Affe, Ratte, Maus, Hund, Rind, Kaninchen, Huhn und Hefe). Mit Ausnahme von Hefe, Huhn und Mensch wurde die DNA aus Nierengewebe isoliert. Die humane genomische DNA wurde aus Plazenta isoliert und die genomische DNA aus Huhn wurde aus Lebergewebe aufgereinigt. Im Autoradiogramm in Fig. 1 wurde als radioaktiv-markierte Sonde ein etwa 720 bp langes hTC-cDNA Fragment, isoliert aus der hTC cDNA, 25 Variante Del2 (Position 1685 bis 2349 plus 2531 bis 2590 der Fig. 6 [Deletion 2; vergl. Beispiel 5 der Fig. 8]), eingesetzt. Die experimentellen Bedingungen für die Hybridisierung und die Waschschritte des Blots erfolgten in Anlehnung an Ausubel *et al.* (1987).

30

Im Fall der humanen DNA erkennt die Sonde zwei spezifische DNA-Fragmente. Das kleinere, etwa 1,5 bis 1,8 kb lange Eco RI-Fragment geht wahrscheinlich auf zwei

Eco RI-Schnittstellen in einem Intron der ghTC-DNA zurück. Aufgrund dieses Ergebnisses ist davon auszugehen, daß nur ein singuläres ghTC-Gen im menschlichen Genom vorliegt.

5      Beispiel 2

Zur Isolierung der 5' flankierenden hTC-Gensequenz wurden ca  $1,5 \times 10^6$  Phagen einer humanen genomischen Plazenta-Genbibliothek (EMBL 3 SP6/T7 der Firma Clontech, Bestellnummer HL1067j) auf Nitrozellulosefilter (0,45 µm; Fa. Schleicher und Schuell) nach Angaben des Herstellers mit einem radioaktiv markierten, etwa 10 500 bp langen 5'-hTC-cDNA Fragment (Position 839 bis 1345 der Fig. 6) hybridisiert. Die Nitrozellulosefilter wurden zunächst in 2 x SSC (0,3 M NaCl; 0,5 M Tris-HCl, pH 8,0) und anschließend in einer Prähybridisierungslösung (50 % Formamid; 5 x SSPE, pH 7,4; 5 x Denhards-Lösung; 0,25 % SDS; 100 µg/ml 15 Heringsperma-DNA) zwei Stunden bei 42°C inkubiert. Für die Hybridisierung über Nacht wurde die Prähybridisierungslösung mit  $1,5 \times 10^6$  cpm/ml Lösung denaturierter, radioaktiv markierter Probe ergänzt. Unspezifisch gebundene, radioaktive DNA wurde unter stringenten Bedingungen, d.h. durch drei fünfminütige Waschschritte mit 2 x SSC; 0,1 % SDS bei 55 bis 65 °C entfernt. Die Auswertung erfolgte 20 durch Autoradiographie der Filter.

Die in dieser Primäruntersuchung identifizierten Phagenklone wurden aufgereinigt Ausubel *et al.* (1987). In weitergehenden Analysen stellte sich ein Phagenklon P12 als potentiell positiv heraus. Eine λ-DNA Präparation dieses Phagens Ausubel *et al.* 25 (1987) und der nachfolgende Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, zeigte, daß dieser Phagenklon ein ca. 15 kb Insert im Vektor enthält (Fig. 2).

Zur Isolierung der vollständigen hTC-Gensequenz wurden in unabhängigen 30 Experimenten jeweils 1 bis  $1,5 \times 10^6$  Phagen mit jeweils verschiedenen radioaktiv markierten Sonden wie oben beschrieben durchmusterst.

Die in diesen Primäruntersuchungen identifizierten, für die entsprechenden Sonden positiven Phagenklone wurden aufgereinigt. Der Phagenklon P17 wurde mit einem etwa 250 bp langen hTC-cDNA Fragment (Position 1787 bis 2040 der Fig. 6) 5 gefunden. Der Phagenklon P2 wurde mit einem etwa 740 bp langen hTC-cDNA Fragment (Position 1685 bis 2349 plus 2531 bis 2607 der Fig. 6 [Deletion 2; vergl. Beispiel 5]) identifiziert. Die Phagenklone P3 und P5 wurden mit einem 420 bp langen 3' hTC-cDNA Fragment (Position 3047 bis 3470 der Fig. 6) gefunden. Nach 10  $\lambda$ -DNA Präparation dieser Phagen und nachfolgendem Restriktionsverdau mit Enzymen, die das genomische Insert in Fragmenten freisetzen, wurden die Inserts in Plasmide umkloniert (Beispiel 4).

### Beispiel 3

15 Um zu untersuchen, ob auch das 5'-Ende der hTC-cDNA im Insert des rekombinanten Phagenklons P12 vorliegt, wurde  $\lambda$ -DNA dieses Klons in einer Southern Blot Analyse mit einem radioaktiv markierten etwa 440 bp langen hTC-cDNA Fragment (Position 1 bis 440 der Fig. 6) aus dem extremen 5'-Bereich hybridisiert (Fig. 3).  
20 Da die isolierte  $\lambda$ -DNA des positiven Klons auch mit dem extremen 5'-Ende der hTC-cDNA hybridisiert, enthält dieser Phage wahrscheinlich auch den das ATG-Startcodon flankierenden 5'-Sequenzbereich.

### Beispiel 4

25 Um das gesamte 15 kb lange Insert des positiven Phagenklons P12 in Teilfragmenten umzuklonieren und anschließend zu sequenzieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einem das gesamte Insert aus EMBL3 Sp6/T7 freisetzen (vgl. Beispiel 2) und zusätzlich im Insert schneiden.  
30

Insgesamt wurden ein etwa 8,3 und ein etwa 6,5 kb langes Xho I-Subfragment sowie ein etwa 8,5, ein etwa 3,5 und ein etwa 3 kb langes Sac I-Teilfragment in den Vektor pBluescript KS(+) (Fa. Stratagene) umkloniert. Durch Sequenzanalyse dieser Fragmente wurde die Nukleotidsequenz von 5123 bp 5'-flankierenden des ghTC-  
5 Genbereichs, ausgehend vom ATG-Startcodon bestimmt (Fig. 4; entsprechend SEQ ID NO 1). In der Fig. 4 sind die ersten (ausgehend vom ATG-Startcodon) 5123 bp dargestellt. In der Fig. 10 (entsprechend SEQ ID NO 3) die gesamte klonierte 5' Sequenz.

10 Um das gesamte ca. 14,6 kb große Insert des Phagenklons P17 in Teilfragmenten umzuklonieren, wurden zum DNA-Verdau Restriktionsendonukleasen ausgewählt, die zum einen das gesamte Insert aus EMLB3 Sp6/T7 freisetzen und zusätzlich einige Male im Insert schneiden. Durch Kombinationsverdau mit den Enzymen Xhol und BamHI wurden ein 7,1 kb, ein 4,2 kb und ein 1,5 kb großes Xhol-BamHI-  
15 Fragment sowie ein 1,8 kb großes BamHI-Fragment subkloniert. Der Kombinations- Restriktionsverdau mit den Enzymen Xhol und XbaI führte zur Klonierung von einem 6,5 kb großen Xhol-XbaI-Fragment, einem 6,5 kb und einem 1,5 kb großem Xhol-Fragment.

20 Die Umklonierung des ca. 17,9 kb großem Inserts des Phagenklons P2 in Subfragmente erfolgte durch Verdau mit dem Restriktionsenzym Xhol. Insgesamt wurde ein 7,5 kb, ein 6,4 kb sowie ein 1,6 kb langes Xhol-Subfragment kloniert. Durch Verdau mit dem Restriktionsenzym SacI wurde zusätzlich ein 4,8 kb, ein 3 kb, ein 2 kb sowie ein 1,8 kb großes SacI-Fragment subkloniert.

25 Das ca. 13,5 kb große Insert des Phagenklons P3 wurde durch Verdau mit den Restriktionsenzymen SacI bzw. Xhol subkloniert. Dabei wurden ein 3,2 kb, ein 2 kb, ein 0,9 kb, ein 0,8 kb, ein 0,65 kb und ein 0,5 kb langes SacI-Subfragment sowie ein 6,5 kb und ein 4,3 kb langes Xhol-Subfragment erhalten.

Die Subklonierung des ca. 13,2 kb großen Inserts des Phagenkrons P5 erfolgte durch Verdau mit den Restriktionsenzymen SacI bzw. Xhol. Insgesamt wurden SacI-Fragmente von 6,5 kb, 3,3 kb, 3,2 kb, 0,8 kb und 0,3 kb Größe sowie Xhol-Fragmente von 7 kb und 3,2 kb Größe subkloniert.

5

Zur Klonierung des 3' von Phagenkron P17 und 5' von Phagenkron P2 gelegenen hTC-genomischen Sequenzbereichs wurden 3 Genomic Walkings mit Hilfe des GenomeWalker™ Kits der Firma Clontech (Katalognummer K1803-1) und verschiedenen Primerkombinationen durchgeführt. In einem Endvolumen von 50 µl wurde 1 µl humaner GenomeWalker Library HDL (Fa. Clontech) mit 10 pmol dNTP-Mix versetzt und in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase Mix (Fa. Clontech) eine PCR-Reaktion durchgeführt. Als Primer wurden 10 pmol eines internen genspezifischen Primers sowie 10 pmol des Adaptor Primers API (5'-GTAATACGACTCACTATAGGGC-3'; Fa. Clontech) zugefügt.

10

Die PCR wurde als Touchdown-PCR in 3 Schritten durchgeführt. Zunächst wurde über 7 Zyklen für 20 sec bei 94°C denaturiert und anschließend für 4 min bei 72°C die Primer angelagert und die DNA-Kette verlängert. Es folgten 37 Zyklen bei denen für 20 sec die DNA bei 94°C denaturiert wurde, die anschließende Primerverlängerung aber für 4 min bei 67°C erfolgte. Abschließend folgte eine Kettenverlängerung für 4 min bei 67°C. Im Anschluß an diese erste PCR wurde das PCR-Produkt 1:50 verdünnt. Ein µl dieser Verdünnung wurde in einer zweiten „nested“ PCR zusammen mit 10 pmol dNTP-Mix in 1xKlen Taq PCR-Reaktionspuffer und 1xAdvantage Klen Taq Polymerase-Mix sowie 10 pmol eines „nested“ genspezifischen Primers und 10 pmol des „nested“ Marathon Adaptor Primers AP2 (5'-ACTATAAGGGCACGCGTGGT-3'; Fa. Clontech) eingesetzt. Die PCR-Bedingungen entsprachen den in der ersten PCR gewählten Parametern. Als einzige Ausnahme wurden im ersten PCR-Schritt statt 7 Zyklen nur 5 Zyklen gewählt und im zweiten PCR-Schritt statt 37 Zyklen nur 24 Zyklen durchlaufen. Produkte dieser Nested-GenomicWalking-PCR wurden in den TA-Cloning Vektor pCRII der Fa. InVitrogen kloniert.

15

20

25

30

Im ersten Genomic Walking wurde der genspezifische Primer C3K2-GSP1 (5'-GACGTGGCTCTTGAAGGCCTT-3') sowie der „nested“ genspezifische Primer C3K2-GSP2 (5'-GCCTTCTGGACCACGGCATACC-3') zusammen mit der HDL-  
Library 4 eingesetzt und ein 1639 bp langes PCR-Fragment erhalten. Im zweiten  
5 Genomic Walking wurde mit dem genspezifischen Primer C3F2 (5'-CGTAGTTGAGCACGCTGAACAGTG-3') und dem „nested“ genspezifischen  
Primer C3F (5'-CCTTCACCCTCGAGGTGAGACGCT-3') aus der HDL-Library 4  
ein PCR-Fragment von 685 bp Länge amplifiziert. Der dritte Genomic Walking  
Ansatz führte unter Einsatz des genspezifischen Primers DEL5-GSP1 (5'-  
10 GGTGGATGTGACGGGCGCGTACG-3') und des „nested“ genspezifischen  
Primers C5K-GSP1 (5'-GGTATGCCGTGGTCCAGAAGGC-3') zur Klonierung  
eines 924 bp PCR-Fragments aus der HDL-Library 1. Insgesamt wurden durch dieses  
Genomic Walking-Verfahren 2100 bp der 3' von Phagenklon P17 gelegenen  
genomischen hTC-Region identifiziert (s. Fig. 7).

15 Die subklonierten Fragmente sowie die Genomic Walking-Produkte wurden einzelnsträngig sequenziert. Unter Verwendung der Lasergene Biocomputing Software  
(DNASTAR Inc. Madison, Wisconsin, USA) wurden überlappende Bereiche  
identifiziert und Contigs gebildet. Insgesamt wurden aus den gesammelten  
20 Sequenzen der Phagenklone P12, P17, P2, P3 und P5 sowie den Sequenzdaten aus  
dem Genomic Walking 2 große Contigs zusammengestellt. Contig 1 besteht aus  
Sequenzdaten von Phagenklon P12, P17 und den Sequenzdaten aus dem Genomic  
Walking. Contig 2 wurde aus den Sequenzen von Phagenklon P2, P3 und P5  
zusammengesetzt. Überlappende Phagenklonbereiche sind in Fig. 7 schematisch  
25 dargestellt. Die Sequenzdaten der 2 Contigs sind nachfolgend dargestellt. Das ATG  
Startcodon in Contig 1 ist unterstrichen. Das TGA Stopcodon ist in Contig 2  
unterstrichen.

## Contig1:

5	ACTTGAGCCC	AAGAGTTCAA	GGCTACGGTG	AGCCATGATT	GCAACACCAC	ACGCCAGCCT	TGGTGACAGA	70
	ATGAGACCT	GTCTCAAAAA	AAAAAAAAAA	AATTGAAATA	ATATAAAGCA	TCTTCTCTGG	CCACAGTGG	140
	ACAAAACCG	AAATCAACAA	CAAGAGGAAT	TTTGAAGAAT	ATACAAACAC	ATGAAATTAA	AAATGATAAC	210
	TTCTGAATGA	CCAGTGGATC	AATGAGAAA	TTAAAAGGA	AATTGAAAAA	TTTATTAAAG	CAAATGATAA	280
	CGGAAACATA	ACCTCTCAA	ACCCACGGTA	TACAGCAAAA	GCAGTGCTAA	GAAGGAAGTT	TATAGCTATA	350
	AGCAGCTACA	TCAAAAAGT	AGAAAAGCCA	GGCGCAGTGG	CTCATGCCCTG	TAATCCCAGC	ACTTTGGGAG	420
10	GCCAAGGCGG	GCAGATGCC	TGAGGTCAGG	AGTTGAGAC	CAGCTGACC	AAACACAGAGA	AACTTTGTCG	490
	CTACTAAAAA	TACAAAATTA	GCTGGGCAATG	GTGGCAGATG	CCTGTAATCC	CAGCTACTCG	GGAGGGCTGAG	560
	GCAGGATAAC	CGCTTGAACC	CAGGAGGTGG	AGGTTGCGGT	GAGCGGGAT	TGGCCATTG	GACTCCAGC	630
	TGGGTAACAA	GAGTGAAACC	CTGTCAGAAG	AAAAAAAAGA	AAGTAGAAAA	ACTTAAAAT	ACACCTAAT	700
	GATGCACTT	AAAGAACATG	AAAAGCAAGA	GCRAAACTAA	CTTAAATTAG	GTAAGAGAAA	AAAGATAATA	770
	AAAGATCAGAG	CAGAAAATAA	TGAAACTGAA	AGATAACAA	ACAAAAGATC	AAACAAATTAA	AAAGTTGTTT	840
15	TTTGAAAAG	ATAAACAAA	TTGACAAACC	TTTGGCCAGA	CTAAGAAAAA	AGGAAAGAAG	ACCTTAAATAA	910
	ATAAAAGTCAG	AGATGAAAAG	AGAGACATTA	CAACTGATC	CACAGAAATT	CAAGGATCA	CTAGAGGCTA	980
	CTATGACCAA	CTGTACACTA	ATAAATGAA	AAACCTAGAA	AAATAGATA	AATTCTAGA	TGCTACAAAC	1050
	CTACCAAGAT	TGAACCATGA	AGAAAATCAA	AGCCCAAACA	GACCAATAAC	AATAATGGGA	TTAAAGCCAT	1120
	AAATAAAAGT	CTCTAGCA	AGAGAAGCCC	AGGCCAACT	GCTTCTCTG	CTGAGTTTA	CCAACTATT	1190
20	AAAGAACAT	GAATTCAT	CCTACTCAA	CTATTCTGA	AAATAGAGGA	AAGAATACTT	CCAAACTCAT	1260
	TCTACATGGC	CACTATTAC	CTGATTCAA	AACCGACAA	AAACACATCA	AAACAAACAA	AAACAAAAAA	1330
	CAGAAAGAAA	GAAAACACTA	GGCCAAATCT	CTCTGATGA	ACTGATACAA	AAATCCTCAA	CAAACACACTA	1400
	GCACACAAA	TAAACAAACA	CTTCGAAAG	ATCTCATT	TGATCAAGT	GGGATTATT	CCAGGGATGG	1470
25	AAGGATGGTT	CAACATATGC	AAATCAATCA	ATGTGATACA	TCATCCCAAC	AAAATGAAGT	ACAAAAAACTA	1540
	TATGATTATT	TCATTTATG	CAGAAAAGC	ATTGTATAAA	ATTCATGCC	CTTCATGATA	AAAACCTC	1610
	AAAAACACCG	TATACAAGAA	ACATACAGGC	CAGGCACAGT	GGCTCACACC	TGGCATCCA	GCACCTCTGG	1680
	AGGCCAAGGT	GGGATGATTG	CTTGGGCCCA	GGAGTTGAG	ACTAGCTGG	GCAACAAAAT	GAGACCTGGT	1750
	CTACAAAAAA	CTTTTTAA	AAATTAGGCA	GGCATGATG	CATATGCC	TAGTCCCAGC	TAGTCTGGAG	1820
	GCTGAGGTGG	GAGAATCACT	TAAGCTTAGG	AGTCGAGG	TGAGTGGAGC	CATGAACATG	TCACCTGACT	1890
30	CCAGCCTAGA	CAACAGAAC	AGACCCCCACT	GAATAAGAG	AAGGAGAAGG	AGAAGGGAGA	AGGGAGGGAG	1960
	AAGGGAGGAG	GAGGAGAAGG	AGGGAGTGG	GGAGAAGTG	AAGGGAGG	GGAGGAGAAG	GAGGAGAAG	2030
	AAGAACATCA	TTTCACATCA	ATAAAGCCG	TATATGACG	ACCGGAGTAG	TATTATGAGG	AAAAAACTGAA	2100
	AGCCTTCCCT	CTAACATCTG	GGAAAATGACA	AGGGCCACT	TTCACCACTG	TGATTCAAC	TAGTACTAGA	2170
	AGTCCTAGCT	AGAGAACATC	GATAAGAGAA	AGAAAATAAA	GGCATCCAA	CTGGAAAGGA	AGAAGTCAAA	2240
35	TTATCCTGTT	TGAGATGAT	ATGATCTTAT	ATCTGAAAAA	GACTTAAGAC	ACCACTAAA	AACTATTAGA	2310
	GCTGAAATT	GGTACAGCAG	GATAACAAAT	CAATGTAAC	AAATCTAGT	TATTTCTATA	TTCCAACAGC	2380
	AAACAATCTG	AAAAGAAC	CAAAAGAAC	GCTCACAAA	AAATTAACAA	GCTAGGAAAT	ACCCAAGAA	2450
	GTGAAAGATC	TCTACAAATG	AAACATATAA	ATTTGTATAA	AAGAAATTG	AGAGGGCAC	AAAAAAAGAAA	2520
40	AGATATTCCA	TGTTCATAGA	TTGGAAGAAT	AAATACTGTT	AAAATGTC	TACTACCCAA	AGCAATTTC	2590
	AAATTCAATG	CAATCCCTAT	TAAAATACTA	ATGACGATG	TGACGATTTA	AGAAGAACAA	ATTCTAAGAT	2660
	TTGTCAGAA	CCACAAAGAA	CCCGAACATG	CCAAAGCTAT	CCTGACAAA	AGAACACAAA	CTGGAGCAT	2730
	CACATTACCT	GACTCTAAAT	TAACTACAA	AGCTTATAGA	ACCCCTAACAA	CATGGTACTG	GCATAAAAAAAC	2800
	AGATGAGACA	TGGACAGAG	GAACAGAAC	GAGAATCCAG	AAACAAATTC	ATGCAATCAC	AGTGAACCTA	2870
45	TTTTTGACAA	AGGTGCCAG	ACACATAC	GGGGAAAAGA	TAATCTCTC	AATAATGTT	GCTGGAGGAA	2940
	CTGGATATCC	ATATGAAAAA	TAACAATCT	AGAACATCTG	CTCTCACCT	ATACAAACCC	AGCAATTTC	3010
	GGATGAAAGG	CTTAAATCTA	AAACCTCAA	CTTGCACAT	TCAGCATG	AAACCTCTCA	3080	
	GGACATTGGA	GTGGGCAACG	ACTCTCTGAG	TAATTCTCTG	CAGGCACAGG	CAACCAAAGC	AAAACAGAC	3150
	AAATGGGATC	ATATCAAGTT	AAAAGCTT	TGCCCAGCAA	AGGAACAAATC	CAACAAAGAG	AGAGACAAAC	3220
50	CCACAGAAATG	GGGAGATATA	TTTGCAAC	TTGCAAC	AAACAAATTC	ATGCAATCAC	AGTGAACCTA	3290
	GCTCAACACTA	CTCTTAAAGA	AAAACACCTA	TAAGAGCTG	TTTCACAAAT	AAGCAAAAGA	TCTGGTAGA	3360
	CATTTCCTAA	AAATAAGTCAT	ACAAATGCGA	AAACGGCAT	TGAAAATGTC	CTCAACACCA	CTGATCATCA	3430
	GAGAAATGCA	AAATCAAAATC	ACTATGAGAC	ATCATCTCAT	CCCAGTTAA	ATGGCTTTA	TTCAAAAGAC	3500
	ACGCAATACAA	AAATGCGAGT	GAGGATGTTG	ATAAAAGGAA	ACCCCTGGAC	ACTGTTGGTG	GGATGGAAA	3570
55	TTGCTACAC	TATGGAGAAC	AGTTGAAAG	TTCTCTAA	AAACTAAAAT	AAAGCTTACCA	TACAGCAAT	3640
	CCATTGCTAG	GTATATACAT	CAAAAAGG	ATTCAGTGA	TCAACAGCT	ATCTCCACTC	CCACATTTC	3710
	TGCAGCACTG	TTCTACAGC	CCAAAGTTG	GAACCAACCT	CAGTGTCT	CAACAGACGA	ATGGAAAAG	3780
	AAAATGTGGT	GGCACATACAC	AATGGGATAC	TACCCAGGCC	TAAAAGGAA	TGAGATCTTG	TCAGTTGCAA	3850
	CAGCATGGG	GGCAGCTGGT	AGTATGTTA	GTGAAATAAG	CCAGGCACAG	AAAGCAACAA	TTTTCTGTT	3920
60	CTCCCTTACT	GTGGGAGCA	AAAATAAA	CAATTGACAT	AGAAATAGAG	GAGAATGGTG	TGTTCTAGAGG	3990
	GGTGGGGGAC	AGGGTGAAC	GAGTCACAAA	TAATTTTATG	TATGTTTTA	AATAACTTAA	AGAGTATAAT	4060
	TGGGTGTTT	GTAAACACAA	GAAGGATAAA	ATGCTGAA	GTGCAAGATA	CCCCATTAC	CCTGATGTGA	4130
	TTTATTACAC	TTGTATGCT	GTATCAAAAT	ATCTCATGTA	TGCTATAGAT	ATAAACCTA	CTATATTAA	4200
	AATTAAAATT	TTAATGCGCA	GGCAGCTGG	CTCATGTCG	TAATCCACG	ACTTTGGGAC	GGCGAGGGCG	4270
65	GTGAGTCACC	TGAGGTGAG	AGTTGAAAC	CAGTCGCCC	ACCATGATGA	AACCCCTGCT	CTACTAAAGA	4340
	TACAAAATT	AGCCAGGGT	GGTGGCACAT	ACCTGTAGTC	CCAACACTC	AGGAGGCTGA	GACAGGAGAA	4410
	TTGCTTGAC	CTGGGAGGCG	GAGGTGTCAG	TGAGCCGAGA	TCATGCCACT	GCACCTGCAG	CTGGGTGACA	4480
	GAGCAAGACT	CCATCTCAA	ACAAAACAA	AAAAAGAGA	ATTAACATTG	TAATTTTAT	GTACCTGATA	4550
70	AAATATACAT	CTACTATATT	AGAAGTTAA	AAATAAAACA	ATTATAAAAG	GTAAATTAC	ACTTAATCTA	4620
	AAATAAGAAC	AATGTATG	GGGTTTCTAG	CTCTGAAAGA	AGTAAAAGTT	ATGGCCACG	TGGCAGAAAT	4690
	GTGAGGGAGG	ACAGTGGAA	TTACTGTG	TTAGACGCT	TAACTCTCTG	TAAGTACTT	AATTAAAC	4760
	AAAGACAGGC	TGGGAGAAGT	TAAGAGGCA	TTCTATAAGC	CCTAAACCAA	CTGCTATAAA	TGGTGAAGG	4830
	TAATCTCTAT	TAATTACCAA	TAATTACAGA	TATCTCTAA	ATCGACGCTC	AGAAATGGCA	CGTCTGATCA	4900
	CACCGTCTC	TCATTCAAGG	TGCTTTTTT	CTTGCTGCT	TGGAGATTTC	CGATTGTTG	TTCTGTTT	4970
75	GTAAACTTA	ATCTGTATGA	ATCCGAAAC	AAAAATGGT	GGTGTATTC	TCCAGAAGAA	TTAGAGTAC	5040
	TGGCAGGAG	CAGGGGGCTC	TGTGGACCTG	AGCCACTTC	ATCTCAAGG	GTCTCTGGC	AGACCCAGG	5110

	TGCAAGGCAG	AGGCCTGATG	ACCCGAGGAC	AGGAAAGCTC	GGATGGGAAG	GGGCATGAG	AAGCCTGCCT	5180
	CGTTGGTGA	CAGCGATGA	AGTGCCCTTA	TTTACGCTT	GCAGAGATTG	CTCTGGATAC	CATCTGGAAA	5250
	AGGGGGCCAG	CGGGATGCA	AGGAGTCAGA	AGGCCCTGCG	TCAAACCCAG	CCACAGCAGT	ATGGGCCCA	5320
5	CCCGGGCGT	TGCCAGAGG	AGAGGACTCA	AGGCCACCTCG	AAGTATGGCT	TAATACCTT	TTTCACCTGA	5390
	AGCAGTGACC	AAAGGTATT	CTGAGGGAAAG	CTTGAAGTGTAG	GTGCTTCTT	TAACACAGAA	AGTCATGGAA	5460
	GCACCCCTCT	CAAGGGAAA	CCAGACGCC	GCTCTCGGGT	CATTACCTC	TTCCCTCT	CCCTCTCTG	5530
	CCCTCGCGT	TCTGATCGG	GACAGACTG	CCCCCGTGA	GCTCTCCGA	GCCCGTGTG	AGGACCTCT	5600
10	TGCAAGGGC	TCCACAGACC	CCGGCCCTGG	AGAGGAGGT	CTGAGGCTG	CTTAATAACA	AACCTGGGATG	5670
	TGGCTGGGG	CGGACAGCGA	GGGGGGATT	CAAAAGCTTA	ATTCACAG	TAATTCAC	CTTCCACAT	5740
	CCGAATGGAT	TGGATTTA	TCTTAAATT	TCATCAAATA	ACATTACAGGA	CTGCAGAAAT	5810	
	CCAAAGGGCT	AAACAGGAA	CTGAGCTATG	TTGCGCAAGG	TCCAAGGACT	TAATAACCAT	TTTCAGAGGG	5880
15	ATTTTCGCC	CTAAGTACTT	TTTATTTGGT	TCATAGGCT	GCTCTAGGT	CAAGGGAAA	GTACACGAGG	5950
	AGAGGCTGG	GGCGCAGGC	TATGAGCAGC	CGAGGGCAC	CGGGGAGAGA	GTCGGGGCC	TGGGAGGCTG	6020
	ACAGCAGGAC	CACTGACCGT	CCCTCCCTGG	AGCTGCCACAA	TTGGGCAAC	CGAAGGGGGC	CACGGCTCGT	6090
20	GTGACTCAGG	ACCCCATACC	GGGTTCTCG	GGCCACCCAC	ACTAACCCAG	GAAGTCACGG	AGCTCTAAC	6160
	CCGTGAAAC	GAACATGACC	CTTGCCTCG	TGCTTCCCTG	GGTGGGTCAA	GGGTAAATGAA	GTGGTGTGCA	6230
	GGAAATGGCC	ATGTAATTA	CAGCACTCTG	CTGATGGGGA	CCGGTCTCTT	CATCATTT	CATCTTCACC	6300
25	CCCAGGACT	GAATGATTCC	AGCAACTCT	TCGGGTGTGA	CAAGGACATGA	CAAACACTCG	TACAAACACC	6370
	ACTCTTCTAC	TAGGCCACAA	GAGCACGSC	CAACACCTG	ATATATTAA	AGTCCAGGAG	AGATGAGGCT	6440
	GCTTTCAGCC	ACCAGGCTGG	GGTGACACAA	GGCGCTGAAC	AGCTGTCTT	TCTAGACTAC	TAGACCCCTGG	6510
	CAGGCACTCC	CCCGAGCTTG	ACGGGCTGT	TGCTGTTTCC	CGAGGGCGCC	ATCTGCCCTG	GGAGACTCAGC	6580
30	CTGGGTGCC	ACACTGAGG	CAGCCCTGTC	TCCACACCTC	CGGCCCTCCAG	GCCTCAGCTT	CTTCAGCAGC	6650
	TTCTAACAC	CTGGGTGGC	CGTGTTCAGC	CGCTACTGTC	TCACCTGTG	CACTGTGTC	TGTCCTAGCG	6720
	ACGTAGCTCG	CACGGTTCTC	CCTCACATGG	GGTGTCTGTC	TCCTTCCCA	ACACTCACAT	GCCTTGAAGG	6790
35	GAGGAGATTG	AGACTGGCT	TCGGCCCTCC	AGACTGGCTC	CTCTGAGCCT	GAACCTGGCT	CGTGGCCCC	6860
	CTCTGGTCTC	GGCTGCACGC	TGACCTCTCAT	TTCCAGGGCTC	TCCCGGCTCT	CTGTCATCTC	CCGGGGCCCTG	6930
	CCGGTGTCTT	CTTCTGTTT	TGTCTCTT	TCCACGTC	GTGCGCTG	TCTCTGCCCG	CTAGGGCTC	7000
40	GGGGTTTTA	TAGGCATAGG	ACGGGGGGT	GGTGGGCCAG	GGCGCTCTTG	GGAAATGCAA	CATTGGGTG	7070
	TGAAAGTAGG	ATGTGCCGTC	CTCACCTAGG	TCCACGGGCA	CAGGGCTGGG	GTGGAGCCC	CCGGCACGGGA	7140
	CCGGCCCTTC	CTGCCCCAGC	ACTTCCCTG	CCCCCTCCCT	CTGGACACAA	GTGTGGCAGT	TFCACAAACG	7210
45	ACTAAGCATC	CTCTCCAAA	AAGACCCAGC	ATTGGCACCC	CTGGACATT	CCCCCACAGC	CCTGGGAATT	7280
	CAAGTGACTA	CGCACATCAT	GTACACACTC	CGCGCACCGA	CGGACCCCCG	CTGTTTATT	TTAATAGCTA	7350
	CAAAAGGGG	AAATCCCTGC	AAAAATGTC	TTAACAAAC	TGTTAAACAA	AACGGGTCCA	TCCGCACCGT	7420
50	GGACAGTTCC	TCACAGTAA	GAGGAACATG	CGGTTTATAA	AGCCTGCAGG	CATCTCAAGG	GAATTACGCT	7490
	GAGTAAACAC	TGCAACCTTC	ATGGGATACG	TACCGAACAT	GCTCAAAAAG	AAAGAAATTTC	ACCCCATGGC	7560
	AGGGGGAGTGG	TTAGGGGGT	TAAGGACGGT	GGGGGGCGCA	GCTGGGGCT	ATGCGACGCA	CCTTTTACTA	7630
55	AAGCCAGTTT	CCTGGTCTG	ATGGTATTTG	CTCAGTTATG	GGAGACTAAC	CATAGGGGAG	TGGGGATGGG	7700
	GGAAACGGGGA	GGCTGTGCCA	TCTTGGCCAT	GGCCGGAGT	CTGGTGTGGG	ATAATGCTCT	AGAGATGCC	7770
60	ACCTCTGAT	TCCCCAACAC	CTGTCGACAC	ACCCGGCCG	GGCCCGAGGC	CTTGGCAGGT	GTATCTCGG	7840
	TGAGGCCCT	GAGGTCTGG	ATCTTCGGG	ACTACCTGCA	GGCCCGAAAAA	GTATCCAGG	GGTCTGGGA	7910
	AGAGGGGGGGC	AGGAGGGATCA	GGGGGGGCA	GCCTCAGGAC	GTGGAGGGC	GTCACTGTG	GGCTGAAAGG	7980
65	GGAGGGGGGG	CCTCGAGCC	AGGCGCTCAA	GGGCCCTCCAG	AGCTGGAAA	AAAGGGGGAA	GGGACCCCTC	8050
	ACGGAGGCTG	CAGCAGGAAG	GCACGGCTGG	CCCTTAGCCC	ACCAGGGCCC	ATCGTGGACC	TCCGGCTCC	8120
	GTGCCATAGG	AGGGGACTCG	CGCTGCCCTC	CTAGCATGA	GTGTTGGGG	ATTGGAGAAC	GCAACAGGAA	8190
70	ACCCATCCAC	TGTGAACTA	GGATTATTTC	AAAAACAAAGG	TTTACGAAA	CATCCAAAGG	CAGGGCTGAA	8260
	GTGCTCCGG	GCAAGGGCAG	GGCAGGCACG	AGTGATTAA	TTTAGCTATT	TTATTTATT	TACTTACTT	8330
	CTGAGACAGA	GTATGCTCT	TTGTTGCCAG	GGCTGGAGTC	TCTGGCTCA	CTGCAACCTC	8400	
75	GGTCTCTG	TTCTGAGCAA	TTCTGCTGCC	TCAGCGTCCC	AAAGTGTG	GATTTCAGG	GTGACCCACC	8470
	ACACCCGGCT	RATTGTTAT	TTTTAGTAA	GATGGGCTTT	CACCATGTT	GTCAAGTGA	TCTCAAAATC	8540
	CTGACCTCTAG	GTGATGCCG	CACCTCAGCC	TCCCAAAAGT	CTGGGATTAC	AGGCATGAGC	CACTGCACCT	8610
	GGGCTTATTA	ACATTCTAA	AACTCTCTG	GGCTCAAGTC	GTAAAGGAGT	CATGGAGTTC	8680	
	AAATTCCCT	TTACTCAGG	GTACCCCTCC	TTTGATATT	TCTGTAATT	TCGCTAGACT	GGGGATAACAC	8750
	CGCTCTGTA	CATATTCACA	GGTTCTGTA	CCACCATGTTA	TCCCATGGG	CCCACTGCAAG	GGGCAGGCTGG	8820
80	GAGGCTGAGC	GCTTCAGGTC	CGAGTGGGGT	GGCCCATGTC	CACTGAGGAC	CTGTAGTGA	ATTAGGGCGC	8890
	AAAGTGTGGAC	ACTGTCTGA	ATCTCAATGT	CTCAGTGTG	GCTGAAACAT	GTAGAAATT	AACTGCATCC	8960
	CTCTCTACTCT	ACTGGGATTC	AGCCCTCTCC	CTATCCCCC	CCAGGGCCAG	AGGAGTTCTC	CTCACTCCCTG	9030
	TGGAGGAGG	ATGATTAATT	TGTTATTT	CTACTGTTG	TTAAGGACCAA	ACTGAAATCCA	CTGTTTATT	9100
85	TTTGTGTTG	TTTGAGAGGC	GGTTTCACTC	TTGTTGTC	GGCTGGAGGG	AGTGAATGG	GGCGATCTTG	9170
	GCTCTACTGC	GGCTCTGCTC	CCCGAGTTCA	AGTGATCTTC	CTGCTCCAG	CTCCCATTTG	GCTGGGATTA	9240
	GAGGCCCCGG	CCACCATGCCC	CGACTAATT	TTTGTATT	TTAGTAGAGAC	GGGGGGGGT	GGGGTTACCC	9310
90	ATGTTGGCCA	GGCTGGCTC	GAACCTCTGA	CCTCAGATGA	TCCACCTGCG	TCTGCTCT	AAAGTGTG	9380
	GATTACAGT	GTGAGGCCAC	ATGCCAGGT	CAGAACATTAC	TCTGTTTAA	AAACATCTGG	TCTGAGGTTAG	9450
	GAAGCTTACCC	CACACTCAAGT	GTGTTGGGT	TTAAGGACCAA	TTGTTAGAATT	TTTTTTATG	TGTTAGAACAA	9520
95	CTCTTGATGT	TTTACACTGT	GATGACTAAG	ACATCATCAG	CTTTCAAG	ACACACTAAC	TGCACCCATA	9590
	ATACTGGGGT	GTCTCTCTGGG	TATCAGCAAT	CTTCATGAA	TTGGGGAGG	CGTTTCTCG	CCATGCACAT	9660
	GGTGTAAATT	ACTCCAGCAT	AACTCTCTG	TTCCATTCT	TCTCTCCCT	CTTTTAAAT	TGTGTTTCT	9730
100	ATGTTGGCTT	CTCTGAGAG	AACCAGTGA	AGTACAACT	TAACCTTTGT	TGGAACAAAT	TTTCAAACACC	9800
	GCCCCCTTGC	CTCTAGGGCA	GAGACAAATC	ACAAACACAG	CCCTTTAAA	AGGCTTAGGG	ATCACTAAGG	9870
	GGATTCTCTAG	AAAGGGGACCC	TGTAATCTA	AGTATTCTA	AGACGGAGCT	ACCTCCACG	GAGCGTACAA	9940
105	CCCCAGGGAG	GGTGGAGGGC	CTCTTCAAAT	GCTAGCTCA	TAATAAACG	ATTTCCCTC	GGCAGTTTCT	10010
	GAAAGTAGGA	AAAGGTTACAT	TTAAGGTTG	TTTGTGTTAG	ATTTCACTG	TTGCGGACCT	CAGCTACAGC	10080
	ATCCCTGCAA	GGGCTGGGGA	GACCCAGAG	TTCTCGCC	CCTTAGATCA	AAACTTGAGC	ACCCGGAGT	10150
110	CTGGATTCT	GGCAAGTCTC	CAGCTGTCT	GGGGTTGTC	GGGGGGCCCA	GGTCAGGAGG	GGACAGTGG	10220
	CCGTGTGGCT	TCTACTGCTG	GGCTGGAAAGT	GGGGCCCTCT	AGCTCTGCA	TCCGAGGCTT	GGAGCCAGGT	10290
115	CCCTGGACCC	CGAGGCTGCC	CTCCACCTG	TGCGGGGGGG	ATGTCAGGAC	ATGTTGGCT	CATCTGCCAG	10360
	ACAGAGTGC	GGGGCCCAAGG	GTCAAGGCCG	TTGTGGCTGG	TGTGAGGCGC	CCGGTGCACG	GGCAGCAGGA	10430
	GGCCCTGGCT	CCATTCTCCA	CCCTTCTCG	ACGGGACCCG	CCCGGTGGGGT	GATTACAGA	TTTGGGGTGG	10500

5           TTTGCTCATG   GTGGGGACCC   CTCGCCGCCT   GAGAACCTGC   AAAGAGAAAAT   GACGGGCCTG   TGTCAAGGAG 10570  
 CCCAAAGTCGC   GGGGAAAGTGT   TGCAGGGAGG   CACTCCGGGA   GGTCCCGCGT   GCCCGTCCAG 10640  
 GTCTCTGGGT   TCGTCGCCAG   CCGGCTCTAC   GGCGCTCCGT 10710  
 CCGGAGCCCC 10780  
 GGTGCGCCGC 10850  
 GCGGATTCGA 10920  
 GCGGGGGAAAG 10990  
 CCAGGTGATT 11060  
 ACCCGCTCTG 11130  
 GGGTCCCCGG 11200  
 TCGGCGGGCG 11270  
 GCGATGCCGC 11340  
 CGCTGGCAC 11410  
 TTTCTGGCG 11480  
 TCTCTGGCG 11550  
 GACATGGGA 11620  
 GCCCGACTGC 11690  
 ACGGGGCCCG 11760  
 CGACGACTG 11830  
 CTGCTGGAC 11900  
 TGATACAGCT 11970  
 ATCGGAACGG 12040  
 AGGAGGGCGG 12110  
 AGCGGGAGGC 12180  
 TGTTTCTGT 12250  
 ACGGGCACT 12320  
 GTCCCTGGGA 12390  
 GGACGAGCT 12460  
 GAGACCATCT 12530  
 GCTACTGGCA 12600  
 CCTCAAGACG 12670  
 CAGGGCTCTG 12740  
 ACAGCAGCCC 12810  
 GGGCTCCAGG 12880  
 AAGCTGGGCT 12950  
 GTGAGGAGGT 13020  
 AGGCAGAGCC 13090  
 TGGACACGGT 13160  
 CGTTTGTATG 13230  
 TGGAGCCGGG 13300  
 TTACCTATAA 13370  
 GGGTGGGAGG 13440  
 GTGTGTTGAC 13510  
 TCACCTGAGG 13580  
 40           AAATTAGCTG 13650  
 TGAAACCCAGG 13720  
 GAAACTCTGT 13790  
 ACTGTTCTCC 13860  
 AGATGGCTCC 13930  
 CGTGTCCCCA 14000  
 TGCTCCCAAGG 14070  
 ATGTAAGACT 14140  
 TTATGTTGGG 14210  
 50           AAAAGTCTGC 14280  
 ATCGAACGGC 14350  
 ACCCAGTTTC 14420  
 CCTGTATTG 14490  
 TGGTCCCCGG 14560  
 GGGTGAATGG 14630  
 CCCCCTGCACG 14700  
 TGGTCCCCGG 14770  
 GGGTGAATGG 14840  
 CCCCTGCACG 14910  
 CTGTCCCCGG 14980  
 GGGTGAATGG 15050  
 CCCCCTGCACG 15120  
 CCCATTGCT 15190  
 TCTTGGTCAC 15260  
 GGCACATGCG 15330  
 60           ATGCATGCTG 15400  
 CTGCCACGTG 15470  
 CCCCTCACTT 15540  
 TCACCTTATT 15610  
 ATGACAGTCC 15680  
 GTCTCCGCCA 15750  
 TTCTATCTC 15820  
 70           TCCATTGAT 15890  
 CATGCTTTTC 15890

	CACTTTCAAG	TGTTCTTAAA	ATACTCAA	GTGTTAAC	TCCTTTAAG	TATTCTTATT	CTGTGATTT	15960
	TITCTTGTG	CACGCTGTG	TTGACGTGA	AATCATTG	ATACAGTGA	CTTTAAAGTA	TTCTTTAGCT	16030
	TATTCCTGTG	TTCTTGTG	CAGTGA	TTGACACT	TTTATGTC	AAAGATAGTA	GAGTATCAAG	16100
5	ATACGTAGAG	TATTTAAGT	TATCATTG	TTATGATT	CTAACTCAGT	TGTGAGTGG	TCTGTATAAT	16170
	ACCAATTATT	TGAAGTTGC	GGAGCCTTG	TTTGTGATCT	AGTGTGTC	TGTTTTCCAG	AACTGTCCAT	16240
	TGAAATTG	ACATCCTGTC	AAATAGTGGG	ATGCATGTC	ACTATATCCA	GCTTATTAAG	GTCCAGTCCA	16310
	AAGCTTCTG	CTCTTCTAG	ATGCATGAA	TTCAAGAAG	GAGGCATAC	TCCCTCACCT	GGGGGATGGG	16380
	TCTGTTCAT	TCTTCTCGT	TGGTAGCTT	TATGTGAGGC	ATTGTTAGGT	GCATGCACTG	GGTAGAATT	16450
10	TTATCTTCT	GATGAGTGA	TCTTGTGGG	ACTTCTATG	CTCTAGTAA	CTAGTAA	TTTTTTTAAA	16520
	TTGCTCTTAG	TACTGCCACA	CTGGGCTCT	TTTGTGATTG	ATTTCTG	TGTTGCTGTT	TTCTGCCCTT	16590
	AATTATATA	TATATATATA	TTTTTTTTT	TTTGAGACA	GAGTCTTGG	CTGTCGCCA	GGGTGAGTGC	16660
	AGTGGTGTG	TCACAGGTCA	GTGTA	ACTT	TACCTTCTG	CCTGAGCCG	CAGCCTCCCG	16730
	AGTAGCTGGA	ACTGCAGACA	CGCACCGTA	ACCTGCTGA	ATTTTAAAT	TTTTCTGGA	GACAGGGTCT	16800
15	TGCTGTTG	CCCAGGCTG	TCTCAAACTC	TGGACTCA	GGGATCCATC	TACCTCGGCT	TCCCAAAGTG	16870
	CTGAATTACA	GGCATGAGCC	ACCATGTC	GCCTA	ATACACTT	ATATCTTAT	AGTGTGGTA	16940
	TGTCCTGTTA	ACAGCATGTA	GTTGAGATT	CAATCCAGTC	TGACAGTC	TGTTTAAC	GATAACCTGA	17010
	TTTATTTCA	TTTTTTGTC	ACTAGAGACC	GGCCTGTTG	ACCTCTGATT	TCCACTTGCC	TGTTGCACTG	17080
	CCCTGTTCCC	TTGTTTCTCA	CCACCTCTG	GGTGTG	TGCGTTTCT	GCCGAGTGTG	TGTTGATCCT	17150
20	CTCGTTGCT	CTCGTCACT	GGGCATTG	TTTATTCT	TTTGTGATTG	TGTTGACCCC	TGATCTTTT	17220
	ATTGTCGTG	TTGTTTTG	TTTATTGAGA	CACTCTCACT	CTGTCACCA	GGGTGGAGTG	TAATGGCACA	17290
	ATCTCGCTC	ACTGCAACCT	CTGCTCTC	GTTCA	CTCTCATC	CTCAACCTCA	TGAGTAGCTG	17360
	GGATTACAGG	GGCCACAC	CACGCC	TGGG	TAATTGTTG	ATTTTAA	GAGATAGGCT	17430
	TGGCCAGGCT	GGTCTCAAC	TCTGACCTC	AAAGTGA	TGCTGCTT	CCTCCACAG	TGCTGGGATT	17500
25	ACAGGTGCAA	GCCACCGTGC	CGGGCATACC	TTGATCTT	AAATAGAAT	CTGAACAT	GCTACCTT	17570
	TCTTGAGCAA	TAAGACCCCT	AGTGTATT	AGCTCTGGC	ACCCCCAGC	CTGTCGCTG	TTTCCCCTG	17640
	TGACTTACTG	CTATCTCAGG	CATCTGACA	CCCCCAAC	CTGAGCTTAA	TAAATATTG	TTCCGTGTT	17710
	GAGTGTGTTCT	GTAGCTTGC	TTGGG	TTTGTCTC	TTTGTCTC	GTCTGCTTC	TGTCCTAGGC	17780
	CCGGCGTCTG	GGGTCCCTT	CTTGTCTT	TGGGTGTT	TCTGTCTT	TTATGCTG	TAACCCCG	17850
30	CTTACCTG	GCTGGCTCC	ATGGCATCTA	GGGACGTC	GGGACCTCTG	CTTATGATGC	ACAGATGAA	17920
	ATGTGGAGAC	TCACAGGAG	GGCGGTC	TTGGG	AGTGTG	GCACCCAGTG	GGCACGTT	17990
	CTTAGCCAGT	GAGTGCAGC	AACTGCTG	CGGCCTGGG	TAGCCTGGA	AAACCCAGG	CATGTCGGG	18060
	TCTGTTGGCT	CGCGGTGTC	GAGTTGAAA	TCCGCAAAC	CTGGGTG	GCGCAGGCT	TGACGGTGT	18130
	GCCTGGGG	GGAGTGTCTC	CTTCAGCTC	TCTGCTGGG	ACCCAGACA	AGGATGAGG	CTCCGAGCCG	18200
35	TTGTCGCCA	ACAGGAGCAT	GACGTGAGC	ATGTTGAA	TTTTAAAT	TCTAGCTG	GGCGGGTGGC	18270
	TCACGGCTG	AATCCCAAGCA	CTTTGGAGG	CCAAGGCGG	TGGATCA	GGTCAGGAGG	TGAGACCCAT	18340
	CCTGGCCAA	ATGATGAAAC	CCCATGTA	CTAAACAC	CTTAAATAGC	TGGGGTGTG	GGCGGGTGC	18410
	TGTAATCCA	GCTACTCGG	AGGCTGAGG	AGGAGAATG	CTTGAACCTG	GGAGTTGGA	TGTCAGTGA	18480
	GGCCGACATTG	CACCACTGCA	CTCCAGCTG	GCAACACAG	GAGACTCTG	CTCAAAA	AAAAAAA	18550
40	AAAAAA	AATTCTAGA	GCCACATTAA	AAAGTAAA	AAAGAAGG	GAAATTAA	TAATAATAGA	18620
	TTTACTGAA	GGCCACCATG	TCCACACCTC	ATCATT	TTGTTATTG	GTGGGAGCAT	CACTCACAGG	18690
	ACATTGACA	TTTTTGAGC	TTTGTCTG	GGATCCCG	TGTA	GGTCTGGCC	ATCTGGGCT	18760
	GGACCTGCTG	GGCTCCCAT	GGCCATGG	TTG	TACCG	ATGTTG	TCCGGGATGA	18830
	CCCTCAGTGA	GCTGGATGTC	CACTGCTGG	ATGTTG	TGCTG	GGTGGGATGA	GGTCGCCAGG	18900
45	AGCTGGATG	GTGTTGTC	GATGGTGTG	GTGAGGTTG	AGGTCCTC	GGCCCTGGG	AGCTGGAGGT	18970
	ATGGAGTCCC	GATGATGCA	GTCCGGGG	AGGTGCGG	TGCGAGGTTG	GGCTGGATG	TGTTGGTGT	19040
	GGATGGTGA	GGTCAGGGT	AGGCTCTCA	GGCCCTCG	AGCTGGAGG	TATGGGATC	GGATGATGCA	19110
	GGTCCGGGG	GGGGTCCGA	GGGCTCTG	TGAGCTG	GGTGTGTT	GGGATGGTGC	AGGTCCTGGG	19180
	TGAGGTCA	AGGCCCTG	GTAGGCTGG	TGTCGGG	CTGGATGTTG	CAGGTC	TGAGGTC	19250
50	GGAGGATG	GGGGTGTG	ATATGCGT	TGAGGTTG	TCCGGATGTTG	GGAGGTCTG	GGTGAGGTG	19320
	CCAGGGCTC	CTGAGTGTG	GATGGGTTG	GTGAGGTTG	TGCGAGGTC	GGTGGGATG	ACCAGGCC	19390
	GGTGTGAGCT	GGATGTTG	TGTCGGAT	GTGAGGTTG	GGGGTGA	GGGGTGGG	CTCTGTTG	19460
	AGCTGGATG	GTGTTGTC	GATGGTGTG	GTCTGGAGT	AGGTGCGG	GGCCCTGG	AGCTGGATG	19530
55	GCAGTGTCA	GATGTTGCA	GTCTGGGG	AGGTCG	AGCCTGCG	GGCTGGG	GGATGATG	19600
	GGATGGTGA	GGTGTGGAG	GAGGTC	GGCCCTCG	GAGGTC	TATGGAGT	GGATGATG	19670
	GGTCCGGGG	GAGGTC	GACCCCTG	TGAGCTG	GTGCGG	TGGATGGTAC	AGGTCCTGG	19740
	TGAGGTGCA	AGGACCTG	GTGAGCTG	TATGGG	TGGGATG	GGGGTCA	GTGAGGTC	19810
	CAGGGCTC	GTGAGTGTG	GTGAGGAT	CGGATG	CAGGTC	GGGGTGGG	GGGGTGGC	19880
	TGTGA	ACTGG	TCTGAGT	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	19950
60	AGGTGATGAG	TGGGATG	GCAGGTC	GGGTTG	GGGGTGGG	GGGGTGGG	GGGGTGGG	20020
	GTCTGGATG	TGGGATG	GGGTG	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	20090
	TGCAAGGTC	GGGTGAGGT	GGCAAGG	GGGTGAGGT	GGATGTC	TATGGGATG	GTGAGTCCG	20160
	GGGTGAGGT	GGGGTGTG	GGATGTC	GGATGTC	GGATGTC	TATGGGATG	GTGAGTCCG	20230
	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	20300
65	CTGGGTGAGC	TGGGATG	TGGGATG	TGGGATG	TGGGATG	TGGGATG	TGGGATG	20370
	TGGTGGGCTG	GATGTC	TGCGGATG	TGCGGATG	TGGGATG	TGGGATG	TGGGATG	20440
	GAATGTC	TGCTG	TGCAAGGTC	TGCAAGGTC	TGGGAGGT	TGGGAGGT	TGGGAGGT	20510
	GTCCGGATG	TGCGGATG	TGCGGATG	TGCGGATG	TGCGGATG	TGCGGATG	TGCGGATG	20580
	GTGAGGTG	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	20650
70	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	20720
	GTCAACCAGG	CCTGCGGTTA	GCTGGATG	CGGTGTC	ATGGTCA	TGGGATG	TGGGATG	20790
	CCCTGCTG	AGCTGGATG	GCTGATC	GGTGTG	GGTGTG	TGGGATG	TGGGATG	20860
	GAGCTGGATG	TGCTGATC	GGATGTC	GGTGTG	GGTGTG	TGGGATG	TGGGATG	20930
	ATGCGGTGTC	GGATGTC	GGTGTG	GGGATG	GGGATG	TGAGCTG	TGAGCTG	21000
	CGGATGTC	GGGGTGGG	GGGGTGGG	GGGGTGGG	GGGGTGGG	TGAGCTG	TGAGCTG	21070
75	CAGGGCTG	GGGAGGTC	CAGGCC	GGGGTGGG	GGGGTGGG	TGAGCTG	TGAGCTG	21140
	CGTGAAGGTC	CCAGGGCTG	CGGTGAGCTG	GATGTC	GATGTC	TGAGGATG	TGAGGATG	21210
	CGCAGGCC	GGGGTGGG	GTATGTC	TGATGTC	TGATGTC	TGAGGATG	TGAGGATG	21280

	TGCGGTGAGC	TGGATGTGTG	GTGCTGGAT	GCTGCAGGTC	CGGGGTGAGT	TGCCAGGCC	CTCGGTGAGC	21350
	TGGATATGCG	GTGTCGGCGT	TCCGAATGG	TGCAGGTCA	GGGTGAGGTC	GCCAGGCCCT	TGGTGGCTG	21420
5	GATGTGCGGT	CTCCGGATGG	TGCAGGTCTG	GGGTGAGGTC	GCCAGGCCCT	TGGTGAGCTG	GATGTGCGGT	21490
	GTCCGGATGG	TGCAGGTCCG	GGGTGAGGTC	ACCAGGCCCT	CGGTGATCTG	GATGTGGCAT	GTCCCTCTCG	21560
	TTAAAGGGGT	TGGCTGTGTT	CCGGCCGAG	GGCACCGCT	GGGTGAGGAG	ATCCTGGCCA	AGTTCCTGCA	21630
	CTGGCTGATG	AGTGTGTA	TCGTCAGCT	GTCAGGCT	TTCTTCTATG	TCACCGAGAC	CACGTTCA	21700
10	AAGAACAGGC	TCTTTCTA	CCGGAAGAGT	GTCAGGAGCA	AGTGCAGAAC	CATTGGAATC	AGGTACTGTA	21770
	TCCACCGCC	TCTGGCTGCG	CCGGAAGAGT	CTTCGGAGCA	CAGCCGGGAC	TCAGCATGCG	CTCTGCTCCA	21840
	CTTGCTGTG	CTTCCCTGGC	TGTCAGCT	TGGCTGGGA	GGCAGGGGCC	CCGTACAGG	CCTGGTCCA	21910
15	GTGGATTCTG	TGCAAGGCTC	TGACTGCTG	GAGCTCACGT	TCTCTTACTT	GTAAATCAG	GAGTTTGTGC	21980
	CAACTGGTGT	CTAGGGTTG	TAAGGCAGAG	GGGATTTAAA	TTAGATGGAA	ACACTACCA	TAGCCTCTT	22050
	GCCCTTCTCC	GGGATGTGGG	TCTGATCTC	TCTCTCTT	TTTCTCTT	TTTGGATGG	AGTCTCACTC	22120
20	TGTTGCCAG	GCTGGAGTGC	AGTGGCATAA	TCTGGCTCA	CTGCAACCTC	CACCTCTGG	GTTTAAAGCGA	22190
	TTCACCGCC	TCAGCTCTC	AAAGTGGCTG	GATTCAGGCC	ACCTGCACCC	ACGGCTGGCT	AATTTTGTA	22260
	CTTAAAGGAG	AGACGGGGTT	TACCATGTT	GGCAGGGCTG	GTCTCGAAC	GTACACTCA	GGTGTACAC	22330
	CCACCTTGGC	CTCCCAAGT	GCTGGGTTA	CAGGCTAACG	CACCGTCCC	AGCCCCCGAT	TCTCTTTAA	22400
	TTCATGCTGT	TCTGTATGAA	TCTCAATCT	ATTGGATTA	GGTCATGAGA	GGATAAAATC	CCACCCACTT	22470
25	GGCAGACTAC	TGCAAGGAGC	ACCTGTCAG	GGAGACCTG	GGGATAGGAG	ATTCACCA	TGAGCTAACT	22540
	TCTAGGTGGC	TGCAATTGAA	TGGCTGTGAG	ATTTTGCTG	CAATGTTGG	CTGATGAGAG	TGTGAGATTG	22610
	TGACAGATTG	AAAGCTGGATT	TGCACTAGT	AGGGACGGGA	GGCCTGGCT	GGGAGATGCC	AGCCTGCTG	22680
	AGCCAGGCC	ATGTTATTAG	CTTCTCTGGT	CTTCCCGAG	GTGACTGTG	GAGGGCTTA	GTCAAGAGAT	22750
	CAGGGCTTCC	CCAGCTCCCC	TGCAACTCG	AGTCCCTGGG	GGGCTCTGTG	ACACCCCATG	CCCCAAATCA	22820
	GGATGCTCG	AGAGGGAGCT	GGCAGCACG	CTCGTCAGAG	GTACATCACG	CTCTGGCTG	GGGACCCCGA	22890
30	CGTGTGCTG	GGGCCATTTC	CTTCGATCTG	GGGGAGGCTC	AGGGCTTCC	CTGTGGAAC	AAGTTAATAC	22960
	ACAATGCACC	TTACTTAGAC	TTTACACGTA	TTAAATGGT	TGCGACCCAA	CATGGTCATT	TGACCAAGT	23030
	TTTGGAAAGA	ATTTAATTG	GGTGAACGGGA	AGGAGCACAG	AGACGTGGT	GTGACCAAGA	TGCTCTTGT	23100
	CACTACTGGG	ACTGGTTGTC	TGCTCTGGGG	GGCTTGGAGG	CCGCTTCTCC	CTGGACAGGG	TACCGTGCCT	23170
	TTTCTACTCT	GCTGGGCTG	CGGCGCTGGG	TCAGGGCACC	AGCTCGGAG	CACCCCGGCC	CCCACTGTCC	23240
	ACGGAGTGGC	AGGCTGTCAG	CCACAGATG	CCAGGCTTAC	GTGTCGGCG	TCCAGCCCC	GTGCCCCCAT	23310
35	GGGGTGTGTT	GGGGGAAAG	GCAAGGCGA	GAGGTGTCAG	GAGACTGTG	GGCTCATGAG	AGCTGATTCT	23380
	GCTCCTTGGC	TGAGCTGCC	TGAGCAGCG	CTCCCGCCCT	CTCCATCTGA	AGGGATGTGG	TCTCTTCTAC	23450
	CTGGCTGGC	TGCGCTGGGG	CAGCGCTGGG	CTACCCCGAGT	GGCTTGTACCA	GAGGGACAGG	CATCCCTGTG	23520
	GGAGGGGCAT	GGGTTCACTG	GGCCCCAGAT	GCACCGCTGG	ACCGAGCTCC	CTGGTGTGA	TGGTGGGACA	23590
40	GTCAACCTGG	GGGGTGAACG	CGGGAACCTGG	CGTCCCCAGG	GTTGACTATA	GGACCAAGGTC	TCCAGGTGCC	23660
	CTGCAAGTAG	AGGGGCTCTC	AGAGGGCTC	GGCTGGCATG	GGTGGCGCTG	TGGCTTCCC	TGAGCTTCCC	23730
	CGTGTGCTG	CGTGGGGTCC	CTAGGCTCTC	ACTGAGTCGG	GGGGGGCTTG	TGGCTTCCC	TGAGCTTCCC	23800
45	CCTAGTGTG	TGCTGTGCTG	AGCAAGCTC	CTGAGGGGCT	CTCTATTGCA	GACAGCACTT	GAAGAGGGTG	23870
	CAGCTGGGGG	AGCTGTGCGA	AGCAGAGGTC	AGGCACTAC	GGGAAGGCC	GGCCGGCTG	CTGACGTCCA	23940
	GACTCCGCTT	CATCCCCAAG	TCTGACGGG	TGCGGGCTG	TGTAACATG	GACTAGTCG	TGGGAGCCAG	24010
50	AACGTTCCGC	AGAGAAAAGA	GGGGTGTGTT	GCTTGGTCTT	AACTTCTT	TTAAACAGAA	GTGCGTTGA	24080
	GCCCCACATT	TGGTATCAGC	TTAGATGAG	GGGGGGAGG	AGGGGGCACG	GGACACAGCC	AGGGCCATGG	24150
	ACGGGGCCA	ACCCATTGTC	GGCAGACAGT	GGTGGCCGGA	GGTGGCGGTG	CCTCCAGAAA	AGCAGCGTGG	24220
	GGGTGTAGGG	GGAGCTCTG	GGGAGGGAGC	AGGGCTCTAG	GACCAACAGA	AGCAGGGGGG	CCAGGGCTG	24290
	GATGCGACG	GGGGCAAGGT	CCTGGATCCG	TGTCTGCTG	TGGTGTACCA	CTTCCGTGCG	CTTCCGCTTA	24360
55	CGGGGGCCCG	GGACGGCC	ACGACTGCCA	GGAGGCCACC	GGGGCTCTGAG	GATCTGGAC	CTTCCCCAC	24430
	GGCTCTGCA	CCCCACCCCT	GTGGCTGGG	TGGCTGGCGT	GACCCGGTCA	TCTGAGGAGA	GTGCGGGT	24500
	AGGTGACAGC	AGGTGTCGG	TGAGGATCC	GTGTCACAA	CACATCGGCC	CAGGAACCCG	TTTCAAACAG	24570
	GGTCTGAGGA	AGTCTGGAGG	GGTTCTAGGT	CCCGGGCTG	GGTGGCTGGG	GACACTGGGG	AGGGGCTGCT	24640
60	TCTCCCTGG	GTCTCTATG	TGGGTGGGG	ACTGGCCGG	TGACTGCC	GGCTCCCTATT	CCCAAGGGAGG	25130
	CCCCGCCAGG	CCGAGCGTCT	CACCTCGAGG	GTGAAGGAC	TGTCAGCGT	GCTCAACTAC	GAGGGGGGCC	24780
	GGGGGGCCGG	CCTCTCTGGG	GCCTCTGTC	TGGGCTTGG	CGATATCCAC	AGGGCCCTGG	GCACCTTCG	24850
	GCTGCTGTG	GGGGGCCAGG	ACCCGCCGCC	TGAGCTGTG	TTGGTACATT	GGGGGGAGG	GGACCCCGGT	24920
	GAGCAGGCC	GCTGGACCTT	GGGAGTGGG	GCCTGATGG	CACCTCATG	TGGGTGGAGG	AGGTACTCT	24990
65	GGGGGGCCCG	CAGGGAGTC	AGGTGACCTT	GTCATGTTG	AGGACACACC	TGGCACCTAG	TGGTGGAGGC	25060
	TTCACTGCTT	CCTGCAAGCA	ATGGGGCCCA	CTGTCACCC	TGACTGCC	GGCTCCCTATT	CCCAAGGGAGG	25200
	GTCACCTG	ATTCAGTTT	CGTCAGAGA	AGGAACCGCA	ACGGCTCAGC	CACCAAGGCC	EGGTGCCCTTG	25270
	CACCCCACTG	CTGACCGAGG	GGTCTCTG	CTTGAGGCTC	AGAGAGGGGA	CACAGCCGCC	CCTGCCCTTG	25270
	GGGTCTGGAG	TGGTGGGGGT	CAGAGAGAGA	GTGGGGGAG	CGGCAAGGG	AGGGCCCTGAC	GGCAGAGGTG	25340
	ATGTCTGAGT	TTCTGCTG	CCACTGTCAG	TCTCTCGCC	TCAACTACACA	CAGGTGGATG	TGACGGGGCC	25410
	GTACGGACACC	ATCCCCAAGG	ACAGGCTC	GGAGGCTCATC	GGCAGCATCA	TCAAACCCCA	GAACACGTAC	25480
	TGCGTGGCTC	GGTATGCCG	GGTCTGGAGA	GGCCGCCATG	GGCACCTG	CAAGGGCTTC	AAAGGCCACG	25550
	TAAGGTTAC	GTGTCAGTGT	GTGTCAGGAG	ATGTGTGCT	CTGGGATATG	AAATGTGCTA	GAATGCGATC	25620
	GTGTCGTGA	TGCTGTTCTG	TGGTGGGGT	ACTTCATGA	TTACACATG	TGTGATATG	TGTCATG	25690
70	GGGTCTGGAG	TGGTGGCTG	TGTCATGATG	TTGGTGGGGT	TTGGGATATG	TGTCATG	TGTCATG	25760
	CCATGGTGTG	TGTGCTG	TGTCATGATG	TGTCATGATG	TGTCATGATG	TGTCATGATG	TGTCATG	25830
	CATGCTGTG	ATGTGCTT	TGTCATGATG	TGTCATGATG	TGTCATGATG	TGTCATGATG	TGTCATG	25900
	GGGTCTGGT	GGGGCTTGGG	CTTACTCTT	CTCTCTCCAG	GCATGGTCCG	CACCATGTC	CTCACGCTCT	25970
	GGGGTCTGG	TTGGGGAGC	TCCACATCA	GGGGCTCTAC	TTCCTAGCATG	GGTGGCCCTG	TCTCTGTCACA	26040
	GGGTGGGCC	TGGGAGACTG	TAAGCAGGT	TTGAGAGGA	AGTAGGGATG	CTGGTGGTAC	CTTCTGGGAC	26110
	CCCTGGCACC	CCCAGGACCC	CACTCTGGCC	TATGCGGCT	CCATGAGATA	TAGGAAGGCT	GATTCAAGGG	26180
	TCGCTCCCCG	GGACACACTC	CTCCCAAGG	GGGGGGGGG	CTGGGGCTC	GGCAGGGGTG	AAAGGGGCC	26250
	TGGGCTGGG	TTCCCAACCA	GTGGTCATGA	GCACGCTGGA	GGGGTAAGCC	CTCAAGGTG	TGCCAGGCCG	26320
	GGGTGAGAG	GTGAAGAAGT	ATCCCTGAGG	CTTCGGCTG	GGGAGAGGCA	CATGTGGAAA	CCCACAGGGA	26390
75	CCTCTTCTC	TGACTCTTG	AGCT					26414

## Contig 2:

5 TGTGGGATTG GTTTCATGT GTGGGATAGG TGGGGATCTG TGGGATTGGT TTTTATGAGT GGGGTAACAC 70  
 AGAGTTCAAG GCGAGCTTC TTCCCTGAGT GGGTCTGCAG GTGCTCCAAAG AGCTTATTG AGGAGACCAT 140  
 ATCTTCCTT GAACATATGGT CGGGTTTATA GTAAAGTCAGG GGTGTGGAGG CCTCCCTG GCTCCCTGT 210  
 CTCTTCTTC CACTCTGGG TCGTGTGGT CCGTGTGTGG TGTGTGGCCCG GTGGCAGGG CTTCCAGGCC 280  
 TCTTGTGTT CATTGGCTG GATGTGGCCC TGCGTACGCT CGCTCCTGG AATTCCTGG CGAGTTGGAG 350  
 GCTTCTTC TTCTTTTT TCTTCTTT TTTTTTTTG TGATAACAGA GTCTCGCTCT TTTTGCCCA 420  
 10 GCGTGGAGTG GTTGGCTG ATCTTGGCTC ATGCAACCT GTGCTTCTG AGTCAAGCA ATTCTCTTGC 490  
 CTCAGCTCC CAACTAGCTG GAATTATAGG CGCCACAC CATGCTGACT AATTTTGTA ATTTTAGTAG 560  
 AGACGAGGT TCTCATGTT GGCCAGGCTG GTCTGAACT CCTGACCTCA GGTGATCCTC CCACCTCGGC 630  
 CTCCCAAGT GCTGGATGA CGGGTGTGAA CGGGCGAGAC TCGCTTCTG CAGCTTCCGT 700  
 GAGATCTGCA GCGATAGCTG CCTGAGCCT TGGTGTGAC AACCTCCGGT TTCTCTCC AGGTCTCCGT 770  
 15 AGGGGCTTT CCATTCATG ACTCTCTCA CAGAAGAGTT TCACGTGTG TGATTTCCCG GCTCTTCC 840  
 GCGTGGATG TGCTGTGTT TTATGGATG CCTCTCTCA TTCTCTTAG GCTTGTGTTA TTGTTGTTT 910  
 TCCGGCTCT TGAAGAAA TTCTGATTA TGATGTTTG AACTTCTTCTT TCTAAACAAAG CATCTGAAGT 980  
 TGCGTTTC CCTCTAAAGC AGGGATCCCG AGGCCCTGG CTGTTGGAGT GCACCGGTCT GGGGCTCTT 1050  
 AGGAAACCCG CGCACAGCGG GAGGGTAGGT GGGGTGTGG GACCGCGGAC TCCGGCTGA GCCCCCGCCC 1120  
 TCTCAGATG CAGTGGCAT CGGGTCTCA GAGGGCACA CACCTACTG AGAACTGTG GTAGAGGGGG 1190  
 20 TCTAGATTCT GTGCTCTTA TGGGAATCTA ATGCTGTGATC ATCTGAGGTG GAAACCGTTG CTCCCAAAC 1260  
 CATCCCTTC CCCACTGCTG TCGTGTGGAA AAATCTCTT CCACGAACCC AGTCCTCTGG ACCACAAATGG 1330  
 TTGGGGACCC TGCTGAAAG CACCTCTCA GCAAGCTCTC TGAGCTGTTG ATATATTGGC TTTCTGTGT 1400  
 TGAGTCAGA ATAATTACGG ATTCTGTGA TGCTTCCCG CGACCTCAGA CCCATGGGC ATTGTGGGGC 1470  
 GTGTTCCGCT CTCCCTGGG TGGGAAGGGT CAGGGCCCAT GTACCTCTCTG GTTACTGTGCT TCCAGGTGG 1540  
 25 TTCTCAGGGT TGAATCGTAC TCGATGTGGT TTAGGCCCCAC CGCCCTCGG CCAAGCTCTG GGGGCTGGGG 1610  
 AACATGCTGA AGCACAGAGT CACCGTGCAC GTCTTTGTAT GCCTCACAAG CTCGAGGGCT CCTGTGTCCG 1680  
 TGTTAGTGTG TGTCACGTG CCTGCTCACAT CGGCTGTGG GAGGAGGAGG GCTTAGCAGG TCCCTCTGG 1750  
 AATGACAGC GCTCTGGGG AGTCTCAGA ATAGGAGGTG GGGGTGCGG TCTCTCTCC GCGCTCTTCA 1820  
 ACTCTCTCC TGCTGTGCT GTGGTGCAC CTGCATCCCTC GCAATCCCTC CAGCACTGGG CTGGAGAGGC 1890  
 30 CCGGGAGCTC CAGTGGCACAT GTGCTGTGAT GGCAGTGGAT GGCAGCTGGT CACGGGGGTG TGAATGTG 1960  
 TGACTGTGGA TGGCGTTGG TACAGGGGT CGTATGTGTG TGACTGTGAT ATGGGGTCTG TGGGGTCTGA 2030  
 TGTTGTGACT GTGGATGGCG GTCGTGGGGT CTGATGTGTG TGACTGTGTTG ATGGGGTCTG TGGGGTCTGA 2100  
 TGTTGTGACT GTGGATGGGG TCGTGTGGGGT CTGATGTGTG GACTGTGGAT GGGGTGCTG GGGTGTGATG 2170  
 35 TGTTGACTGT GGATGGCAGT CGTGGGGTCT GATGTGTGGT GACTGTGGAT GGCCTGCTG GGGTGTGATG 2240  
 TGTTGACTGT GGATGGCAGT CGTGGGGTCT GATGTGTGGT GACTGTGGAT GGCCTGCTG GGGTGTGATG 2310  
 TGTTGTGACT GTGGATGGGG TCGTGTGGGGT CTGATGTGTG TGACTGTGAT ATGGGGTCTG TGGGGTCTGA 2380  
 TGTTGTGTA CTGTGGATGGG CGGTCTGGG TGCTGTGATG TGACTGTGAT ATGGGGTCTG TGGGGTCTGA 2450  
 TGTTGTGTA CTGTGGATGG TGATCGTCA CAGGGGTCTG ATGTGTGTTG ACTGTGGATG GCGGTGCTGG 2520  
 40 GGTCTGTGATG GTGGTACTG TGGATGGTGA TCGGTACAG GGTCTGTGAT TGTTGTGATG GGTACTGTG 2590  
 GTCGTGGGGT CTGATGTGTG TGACTGTGTTG ATGGGGGGT TGCCCGGGGGT TCTGATGTG GGTACTGTG 2660  
 GATGGCGATC GGTCAACAGGG GTCTGTATGT TGTTGACTGT GGATGGCGGT CGTGGGGGTCT GATGTGTGGT 2730  
 GACTGTGGAT GCGGGTCGTC GGGTCTGATG TGTTGACTGTG GGATGGCGGT CGTGGGGGTCT GATGTGTGGT 2800  
 GACTGTGGAT GCGGGTGGT CCGGGGGTC TGATGTGTG TGACTGTGGA TGGGGTCTG GGGGCTGTGAT 2870  
 45 GTGGTACTG TGGATGGCA TGCTGGGGGGT TGATGTGTG TGACTGTGGA TGGGGTCTG GGGGCTGTGAT 2940  
 GTCTGGTGTAC TGCTGGATGGC TGCTGGGGGGT TGATGTGTG TGACTGTGGA TGGGGTCTG GGGGCTGTGAT 3010  
 GTCTGGTGTAC TGCTGGATGGC TGCTGGGGGGT TGATGTGTG TGACTGTGGA TGGGGTCTG GGGGCTGTG 3080  
 ATGCTGGGTG ACTGTGGATC GCGGTCTGG GGTCTGTGAT GGTACTGTG TGACTGTGCTG GTGGGGTCTG 3150  
 ATGCTGGGTG ACTGTGGATC GCGGTCTGG GGTCTGTGAT GATGTGTGGT GACTGTGGAT GGCCTGCTG 3220  
 50 GGTCTGTGATG TGTTGTGACT GTGGATGGGG TGCTGTGGGGT CTGATGTGTG GACTGTGGAT GGCCTGCTG 3290  
 GGTCTGTGATG TGTTGTGACT GTGGATGGGG TGCTGTGGGGT CTGATGTGTG TGACTGTGTTG ATGGCAGTCTG 3360  
 GTCACAGGGG TCTGATGTG TGTTGTGACT GTGGATGGGG TGCTGTGGGGT CTGATGTGTG TGACTGTGATG 3430  
 GGGGTGCTGG GGTCTGTGATG TGCTGGGGGGT TGATGTGTG TGACTGTGGA TGGGGTCTG 3500  
 TGGCGGTCTG GGGGTCTGATG TGTTGTGACTG TGGATGGTGA TGCTGTGACAG GGTCTGTGATG TGTTGTGACTG 3570  
 GCAGGGTGGAG TCCCGAGGTG TGCTGTGACTG ACTTTGGCTC CTGGGGGGGGT TTCCCAAACAA 3640  
 55 GAAGCTCTCC AGGGCGCTCTC TGGGGCTCATC CCGGGCTGGG CGCTGCTGACCA CGCTCTGATC 3710  
 GGAAGAACAA AGTGGCCAGC TCTGGCCGGG GCAGGCCACAA TTGTCGGCTC ATGCCCTCTC CTCTGCGGGC 3780  
 AGGTCTCTCA CTGACAGAC CTCAGCCAGCA ACATGCGACA GTTCGTGGCT CACCTGCGAGG AGACCAAGCCC 3850  
 GCTGAGGGAT GCGGTCTCA TGAGCAGGAGT CTGGGGACTG CGCTGAGGGG TTGGGACGGG ACTCCACAGA 3920  
 GTGGGTCTCTC CCTCTGGGAA TCACTGGCTC CATGACCCAGA CAGACTGTG TGCTGGGGGG GCAAGTGGGG 3990  
 60 GAATGAGCTC TGATGGGGGG TGATGGGGGG TGCTGTGGGGT CGCAATCTG AGCTGTGGCA TGCCAGGCTG 4060  
 CGACAGCTGC TGCTGGCTC CACCTGCTCA CGTGTGACTG CGGGCTGGCT CTCTGATTC CGAGTGGCTT 4130  
 TTTTCATGAT TTGCTTAAATG TCTCTCTGAGT CAGTTTGATG CTTGAGGGCA AAGGAAGGGT GTCCCCCTCC 4200  
 TTTAGGGGGG CAGGGCATGT TTGAGGGCTG TGCTGGGGGGT CGCTGGGGCT ATGTCGGGGT CTGAGGGCAA 4270  
 AGGAAAACGTC TCCCCCTCTC TAGGAGGGAGC GGCGCTGGGGT GAGGACACGGC CGCGTGGGG GGGCTCTCAG 4340  
 65 TGCTGGGTCT GTCCACGTGG CCTCTGTGGGG CTTTGAGATG TGCTGTGCTG CACGTGGCCC TGTTGTCTT 4410  
 TGCGAGATGCC TGTTAGGACT TGCTGGCTC TAGGGGACAG TGCTGTGCTC CGCATGAGGC TCAGAGACCT 4480  
 CTGGGGAAAT TTCTGGCTC CCCAGGGTGG GGGTGGAGGT GGGCTGGGGT GCTGGGACCC AGACCCCTGTG 4550  
 CCCGGCAGCTG GGGCAGCAAC TCCCTGGATCA CATATGCCAT CGGGGGCACC GTGGGGTGTG TGTTGTGAG 4620  
 CCCAGCTGGA CCCACAGGTG GCCCCAGAGA GACGTCTGTG TCACACACT CTGCTCTGAGG CCATGTGTGT 4690  
 CTGCGAGAC TCGGCCCCGG CAGGGCAGGA TGCCCTGCA TTCCAGCCCA GCCCCGCACT TCATCACAAA 4760  
 CACTGACCCC AAAAGGGACG GAGGGTCTTG GCCACGTGTT CTGCTCTGTC TCAGCACCA CCGGGCTCACT 4830  
 CCCATGTGTC TCCCTGGCTC TTGAGGGAGA CTCCCTCCCTG ATGAGGGCA CGAGTGGCTC CTTCGAGCTC 4900  
 TTCTCTACGCT CTATGTGCA CCACGGGGGG CGCATCAGGG GCAAGTGGAGT CAGGTGGCCA GGTGCCATTG 4970  
 CCCTGGGGT GGCTGGGGGG GCTGGCAGGG CCTCTGCTCA CCTCTCTCTC GCCCCCTCCCC CACTGNCCCT 5040

5	CTGCCCCGGG CCACCAAGAGT CTCTTTCTT GGCCCCCGCC CCCTCCGGCT CCTGGGGCTGC AGGCTCCCGA 5110 GCCCGGGGAA ACATGGCTCG GCTTGGCGCA CCCGGAGGGG AGCAGGTGCC ACACGAGGCC TGGAAATGCC 5180 AAGCCGGGTC TGGAGTTGCT CTCGGCTGGA CGACGGAGGG CGGGGGGTGT GCTGGGTCA GTGTGGGCC 5250 GAGCGTTG GCTTCGACGT TGTCACTGC AAGTACTAC TGACCGCTGA CACCCGCTC TCACACGCTT 5320 GTATCTCTCT CTCCCGATAC AAAAGGATT TATCCGATTC TCACTCTGT CCTCTGTGT TGACCCCCGC 5390 GAGGGGGCGG GCTCTCTCTCT CTGTGACTAG ATTCCCATC TGGAAAGTGC GGGGTTGACC GTGTAGTTG 5460 CTCCTCTCGG GGGGCGCTGTG CTGGCGATGG CGCAGCGGC CTGGGAGGC TGCGGTACA CAGGCACTGG 5530 GTGAGCCACA CTACAGGTGG TAGGCCACA CTGGCTGGT CCACATCAGC TCTCTCTGGAT TTTAAGTAAA 5600 ACCCACACCC TCCCGGCAAGG CATCTGGCTG CGACCCCTGTG TGTCCTGGG GAGAGTGGTA GCACGGAGGA 5670 10 AATTCTGGA CACTCAAGGT CATCAGCAAG CTACATCGCA GCTCAGGGC ACCTGGAGGC CTCTCTCTGG 5740 GATCGCTCC AGCGGATAAA GGACTGTGCA CAGCTTCGGA ACCTTTTATT TAAAAAATATA ACTATTAAATT 5810 ATTGCTTATG AGTAATACAC TAATGGTATG ACAGAATTATA ATATTAAATT AGATATAATT AGAAATATTA 5880 AGTAGTACAC ACGTTCTGC AAAACACAAA TTGGCACATGG CAGCAGGTG AATTTCGCC GAGGGACAGC 5950 TGTGACATG TGTGTAAGGG GCCCCCGAGC CCACAGAATT CCTGACGAAAC GTCACCTCCC CAGAGAAGGC 6020 ACCCACGGCC TCCCTGGG TCGTGAATT TATTAGATG GATCACTGA CGTACCGTCC ACCTGGGCA 6090 GGGCTTGGG GAATGTGAGG TGATGACTGC CTGCTCATGC CTGACAGAC AGAGGGTGT TGTGTC 6160 CTGCTCTAG GACACGGACA GGGCGGAAGC TCTAGTCCC ATCGTGTCC AGTTTGGCCT CTGAATAAAA 6230 ACGTCTCAA ACCTGTGCA AAAACAAAAG AAGAACAGAG AGAGTTTCCC ATCCCATGTC CTCAAGGGG 6300 CGTATCTGTG TGCGTGTACT CGCTGGGTG CGCGGACTCC TAGAGTTGTG CGCTGTGCTT CTGTCAAA 6370 20 AGTGCAGTCC TCTTGCCTAT CACTGTGATA CTGCGAACCC CAAGGAAAGC CTCTTTCTT TTCTCTCTT 6440 TTTTTTTTT GAGCGGAAAC GTCACTGTG CTGCTGGG CTGAGTGC CTGGCGGAT CTCAACTCAC 6510 TGCACCTCC GCGCTCGGG TTCCAGCAT TCTCTGCGT CAGCCTCCCG AGCAGCTGAG ATTACAGGCA 6580 CCCACCCCT GCGCTGGCTT ATTCTTGTAT TTTTTAGTAG AGAGGGGTTT TTGCCCCATGTG GGCCAGGCTG 6650 GTCTGAACT CTCGACCTCA GGTGATCCAC CCACCTCGGC CTCCCAAGT GCTGGGATTA CAGGTGTGAG 6720 CCATCACGCC GCGCTGGAAAC GCGCTTTTGTG AGGTGACCA CCTATAGCGC TTCCCGAAA TAACAGGTCT 6790 TGTTTTGCA GTAGGCTGCA AGCGCTCTT GACCAAGAGA GTGGCGTCT GTGGGCTCTG GGGATGGCTG 6860 AGGGTCCGGT GCGACCCATG CTCCTGGTGT GACCTTTAG GTTCCACGGG GCTATTCTGC TCTCACTGT 6930 TGTGTCAAA CGCACCTGTG CGATCTTGTG TTGGAGAGTT TCTGCTCTC TGTCATG CTGAAACTAG 7000 30 GGGCAAGGTT GTATCGTTG GGGCGCAGGG GCTACATGTA GGTCATGAG TCTTTCACCG TGGACAAATT 7070 CCTGAAAAAA AAAAAGAGA TTGGGTTAA GCATTCATTC CGGGTCAAGT GTCTGGTCT CTGAATAAAC 7140 TCTAAGATT AAAAACACCTT ATGAAAGAA AACCTTGATC ATTCAAGGCA AGGATGTGTT CACACCTGTG 7210 GCTGATCTG TTTCAGCCCG CCCAGTGCAT GTTGAGAGTG GGGAGCAGGG ATTGTTTGTG CAGAGGTCTC 7280 ATCTGTTATG TTCTGAGGT TTTCAGGGC TGATGTTAG ACCTGTGCTT TGTCATGATG AGGTTCTGTG 7350 TCTGTTGTTG CTCTGGTTG AGTGTACGCA TGTCAGCAC ATGCCCTGC CGTCTCTCAG CTGTC 7420 35 CGGCCCCCAGG TCTCTAGTCC AGTGGCAGGG GATCCCGAG GGTCTCCATC TCTCCACGCT GCTCTGCAGC 7490 CTGTGCTACG GCGACATGGA GAAACAAGTGT TTGCGGGGA TTGCGGGGA CGGGTGAGGC CTCCCTTCC 7560 CCAGGGGGGTTTGGGGGG TTGGGTTGGG TTGATGTC TTAGGTGTTA ATATTCTGG TGCTCTGGAG 7630 ACCATGACTG CTCTGTCTTG AGGAACCCAGA CAAAGGTGCA GCCCCCTCTT GTGATGAGC CGCACGGGAG 7700 GGTTGCAAC GCGTGGAGAC TGCGGGCTCC AGCGGAGCTC TGTCACGGG CGATCTCAGG AGGCCCTCAGG 7770 40 GCTCAGCAGG CGGGAGGGCC GTCGGCTGC ATGATGAGCA TGTAATTCA ACACGGAGGA AGCACACCG 7840 CTTCTGTCAC GTCACCCAGG TTCCGGTTAGG GTCTCTGGGG AGATGGGGCT GTGTGACACT GAGGGCCAC 7910 ATCTCCCAGG AGGGCCCTGA CAGGTGGCT GGAACGGGGC CCTCTCTCAGC CAATGCCCA TCCCACTTGC 7980 ATGGGGTCTA CACCAAGGA CGCACACACC TAATATCTGT GCAACCTAA TGTTGTCAA CTCACTGGC 8050 45 TTTTATTGAC AGCGATTTACT TTTTTTTTTT TAATACTTA TGTTGCTAGG TACATGTGCA CGACCTGTGCA 8120 GTTAGTTACA TATGTATACA TGTCCTAGT TGTCCTGCTG CACCCATTAAC CTCACTATT ACATTAGGTA 8190 TATCTCTAA TGCTATCCC CCCACCTCC GACAGGCCCT GGTGTTGAT GTTCCCAC 8260 CTGTGTCAA TGTTCTCTAT TGTTGCTTC CCACCTGTG TGAGAACAT GTGTTGTTG TTTTCTTTC 8330 CTTGCAATAG TTGCTCAGA TGATGTTTT CGACTCTGTG CGTCTCTGTA CCATGCTCT ACAAAGGACA TGAACCTCATC 8400 50 CTTTTTATG ACTGCAATGT ATTCCCTGGT GTATATGTG CACATTCTCT TAATTCAGTC TATCATCGAT 8470 GGACATTGGG TTGGTTGCA AGTCTCTGAT CTGTAATGGG TGTCGCAAT AAACATACCT GTGATGTGT 8540 CTTATAGCA CGATGATTA TARTCTTGTG GGATATACCG CAGTAATGGG ATGGCTGGG CAAATGGTAT 8610 TTCTAGTTG ATGATCTGAA GGAATCAACCA CACTGCTTC CACATGGTT GAACTAGTTT ACACCTCCCAC 8680 55 CAACAGTGTAA AAAGTGTCTG GTGTCCTGGAG AGGATGTGGA CAGCGAGTT TTTTTTATGA AAATAGTATC 8750 ACTGCAAGAG CAGACAGTGTG TGAGGAGATG CGTCAGGAAG CGTCAGGCC ACACAGCCAT TTCTCTCGAA 8820 GACTCCGGGT TTCTCTGTG CATCTTGTG AACTCTAGCT CCAATTATAG CTGTAAGT GGATCAAGGT 8890 TCTCTCTCAT TAAGGTTCAA GTCTCTAGTT GAAATAAGTT TATGTAACAG AAACAAATAAT TTCTCTGACA 8960 CACAACCTGC TCTGGGATTG GGAGGAAGT GTCCTCGAGC TGCGGGCACAC CTGGTCAAGCC 9030 GGATACCTCT GGGCCATGGT CATGGGGCCG TGCGCTTGGG CCTGAGGGTC ACACAGTGC CAATGCCAC 9100 60 CTCTCTGTGAT AGTGGATCTG GGTCTCGGAT CATGCTGAGG ACCACAGCTG CCATGCTGGT AAAAGGGCACC 9170 ACGTGGCTCA GAGGGGGCGA GTGTCCTGGAG CGCCAGCTTC TACCTGGCT CAGTTATTTC TCCCTAAAGAG 9240 TCTGAGAAGT GGGGGCCCGC CTGATGGCCT TGCTCTGTCT TCAGCTGGCA CAGAATTGCA CAAGCTGATG 9310 GTAAACACTG AGTACTATTA ATGAGTGGG AATTGGCTGA CGACTTAAGT GTAGAGGAGCT CGTCTGTG 9380 AAAGAAATTG AATTTTCA TTAAACCGCT TTGGAGATG TTAATTTATT TATGGCTGTG TAAATGGTTT 9450 65 GACATTCTG CTCCTCGAGA CAGATACTAC GTAAAAGTG TAAAGTTAAC TTGCTGTGT ATTTTCTCTT 9520 ATTTTACGGT GCTCTCTGGT TTGGTGGATG ATTCTCTGTG GGTGACACTC CACCTCACCC ACACGGAAAAC 9590 CTTCTCTCAGG TGAGGCCGGT GCGCTGGTGC TGTTGGGACC TCCACAGCTT TGCGCTTTG CAGTTGAGGC 9660 CCCCGTGTCC TGCCCCCTGGC ACCGCAGCGT TGCTCTGTCC AAGTCTCTC TCTCTGCCGG TGCTGGATCC 9730 GCAAGAGCAG AGGGGCTGG CGCTGCACCC AGGGCTGGGG CGCGAGGGGC ACCTTCGGGA GGGAGTGGGT 9800 70 ACCGTGCAAGG CCGCTGGCTC CGAGAGACCC ACCCGGTTA CACACGGTTG GAGTGCAGGC GGACCTGG 9870 CTCTCTGTG TCTTGGAAA GTCAAGAGTG GCGCTCCCTG GGCCCCAGT GAGACCCCA GGACCTGTG 9940 ACAGGGCCTG CAGGGCCAGG CGCCAGCGCT CCTCCCGAGG GTGCACTGA GCCTGGGGAG AGCAGGAGCT 10010 GCTGAGTGTGAG CTGGCCACAA CGTTCGCTG CGTTCACGTT CCTGGCTGGG TTGTTGGG ATCGCTGGGA 10080 GAATTGGAT TTGCTGAGTG CTGCTGTCTT GAACCACGG AATGGCTAGG AGTGGTTTC AGAGTTGATT 10150 75 TTTGTAACCTC AAACATAAATG CAGGACACGG GGACCTGGCC TCAGCACAGG GGATTGTC TATGTTGCC 10220 CCTCAAGGGC GCCCCACAGA CGCGTGGGG TTGTTTAAA TGCGATTG ACCAGGGAGC AGAACCTTG 10290 AAAGCTGTAAGGAAACCT CAGAAAATGT GGGCCCAAGG GTGTTTCA GTGTTTGC TGGCTGTGT 10360 TTGTAACCTC CCATTTGGAC CGGCCCTCCA AGTCCACCC CTCCAGGGC CGCCCTGGG 10430
---	---

TGGGGGTATG CCTGGCGTTC CTTGTGCCG AGCCCCGAGC ACAGCAGGCT GTGCACATT AAATCCACTA 10500  
 AGATTCACTC GGGGGGAGCC CAGGTCCAA GCAACTGAGG GCTCAGGAGT CCTGAGGCTG CTGAGGGGAC 10570  
 AGAGCAGACG GGGAACGCTG CTTCGTGTC GCAAGTCCCT GAGGGTCTG GCCAGGGAGG TGGCTCAGAG 10640  
 5 TGTATGTGG GGTCCACCCG GGGGCAGAAC TCTGTCTCG ATGAGTCGGC AGCCATGTAA CAGGAAGGGG 10710  
 TGGCCACAGG GAGCTGGGA TGCAACGAGG GAGCTGGCA GCTGGCCAG GCTCCAGGGC CAGGCCACAG 10780  
 GAAGGGCAGG GGGACCCCG GGGCACACAG AGAGGGCGCA GGAAGGGAGG GGGATGCCCA GGCCAGAGCA 10850  
 GAGGCTACCG GGCACAGGGG GGCTCCCTGA GCTGGTGAAG CGAGGCTCAT GACTGGCCGA GGGAACCTCC 10920  
 TTGACGTGAA GCTGACGACT GTGTTGCCC AGCTCACAGG CCAGGCGAGT CCCGGCCCTG AGCAGAAC 10990  
 CAGAACCTC CCTTGTCT AAACGACAGC AGATGCTTC AGGGGCATCA GGAGAAAACA GGCAAAGTCG 11060  
 10 TTGAGAAACG TCTTAAAGA AGGTGGGATG GTGGCAATT CTTGTCCAGA TTTTAGTCTG CCCCAGGACCA 11130  
 CAGATGACTC TATAACGGG TTGTTGTTG GCACTGGGA CACATGAGAT GGACCATCAG AGAGGGCACT 11200  
 GGGGCTGAC GTCCTCATCG AGTCCCTGCTG TGCCCGGGTC CAGGGCAGGT TCTTGTCTG TCACCTACCT 11270  
 GTCCGTCCCG GGAGACAGGG AAGACACCC GAACTCTGGA GCAGGGCTGG GTCCAGGCTC CTAGAGGCTC 11340  
 CTGCGACGGC CAGCACCCCT CTCACATTA CCACCTCTC GGGGTTTCTT AAAGGATTTA ACAAGGGTGT 11410  
 15 CAGTTTACCT CCTGGGTGAC GGCCCCCATC CCTGGGTGTC AGATGCCCCC TCTGCCCCG GACCCGGTC 11480  
 CGAGGTGTC CTGAGTATGG CTGGTGGTG AACTTGGGA AGACAGTGGT GAACTCCCT GTAGAAGACG 11550  
 AGGGGCTGG TGGCAGCGCT TTGGTTCAGA TGCGGGCCCA CGGGCTATCTC CTCCTGGTGG GCTCTGGTGT 11620  
 GGATACCCGG AACCTGGAGG TGCAAGCCG CTACTCCTGG TGAGCCGACCC TGGGGGAAGA TGAGGCTGT 11690  
 20 GCGGGCTGG GGCAGGTGCT GCTGCAGGGC CGTTGCGTCC ACCTCTGCTT CGGTGTTGGG CAGGGCACTG 11760  
 CCAATCCAA AAGGGTCAGAG GGCACAGGGC GGGCCATGGT CCATCTGGG CTGAGCAGAA ATGCATCTT 11830  
 CTGTTGGAGT GAGGGTGTGTC AACACGGGGAG CAGTTTCTG TGCTTATTTC TGAAAAGGAA ATGGTGCACCA 11900  
 AGACCTGGGT GCACTGAGGT GTCTTACAA AGCAGTCTGG ATCCGAACCC AGACCCCGG GGGCCCTGCTG 11970  
 GGGCTGAGTC TCTCAACCCC GAAACAGGGC GGGCTGTCG GCTGAGTCTG CTCTGAACCC GAGACCTGG 12040  
 25 GGGGCTGTC GGGCTGAGTC TCTCCGAAAC CAGAGACTTC AGGGGCTTGT TGGGGTGAAG TCTCTCCGT 12110  
 GTGAGCCCA CACTCCAAGG CTCATCCACA GTCTACAGGA TGCCATGAGT TCATGATCAC TGTGACCCA 12180  
 TCAGGGGCA GGGCCATGGT GTGTTGGGG TCTCTACAA ATTCTGGGG CTGTTTCTC CAGAGCCCGA 12250  
 GAGCTCAAGG CCCGCTCTA GGCTCAGACAA CAAATGAAAT GAGATGAGAC AGAGATGCAAA 12320  
 TGTTCCTTT ATGAATAAA AGTATCAACA TTCCAGGGAG GGCAGGTGGT CTCACACCTA TAATCCCAAGC 12390  
 ACTTTGGGAG GGGAGGTGG TGTTGATCT TGAGTTAAC GCACTTAAAC GACATAGTGA 12460  
 30 AATTCCTATT CTACTAAAAA AATACAAAAA TTACGCTGTC CTGGTGGCAC AGCGCTGTAG TCCCCCTAT 12530  
 GCGGGAGGCT GAGGCAGGAG AATCATTGGA ACCCAGGAGG CAGAGGTGTC AGTGAGCCGA GTCACACCA 12600  
 CTGCACTCCA GCGCTGGCAA CAGAGTGGG TGCTCATCTT AAAAAAAA AAAAAAGTATC AGCATTCCAA 12670  
 AACCATAGTG GAGGGTGTGTTTATCTC TGCTCTGGAT ATTATTTACT CGTGTGTC TAGAGGCCCC 12740  
 35 AACTGGGGT GCGCTCCCT GAAAGGACAA CCTTCATGGG AAGAGAAAATA AGTGGTGAAT GTTGTAA 12810  
 CCAGAGGTGTT AAACCTGGGTT CCTGCTGTC TGAGTTAAC GTCAGCATG GCACTTAAAC GACATAGTGA 12880  
 ATGCTCCCTG GGGTTTCTGTT CATGGGGAGG CAGCAGGTGT GAGACCCCTC GTGATGGGG AGCAGCAGGT 12950  
 GCGAGCCGCC TCATGATGGG GGAGTGGGAG TGCAAGACAC CCTTGTGCTG GTGCCCCAGC ATGCTCCCTG 13020  
 TGCAGCTCCC GCGCCACAAAG GATGCCACAG TCCCTGTGCTC CCCACAGTCC CTGCTCCCT CTACACGCT 13090  
 TACCTGGTCC TGGCTCCAC TGGCTTGTG TGCTGATTG CCTACATTTCC TGGGGCTCC CAACCTCTC 13160  
 40 GCCTCTCCCA GGCACCTCTG CAGTGTGTCG CATAACGTC AGCTGTGAA TGCTCACTGC TTATTTGCT 13230  
 CCCCATGAAA TGTTTTTCTT AGGACAGGGC CCCCTGGTGC CAGGGCTCTG CAACGATCA GTGAATGTTA 13300  
 TTGAAGGACA AAGGACAGAC AAAACAAATC GAAAGATGGG TTCTCTCTA ACACATGCA AAGCCACAGA 13370  
 GCGTACTGCA GGATGGGGT GGCACTGGTC ATCAGATG TGCTCAATCC CAGAAATATTG TGTCCTCCA 13440  
 45 AAGGGCAACTT GGTCAAGAGT TGTTGCTGCA GAGGGGGCTC TAAAGGCTCA GCACTGGAGG CAGTGGTGG 13510  
 CCATACCTAG GGTGAACACTA CATCTCTGT TGCTGAAGTA TACAGCAGAG CCTGGAAGGG CATCTGGGAG 13580  
 AAGAAACACG GCAAAATGAT TAAGAAAAGT GAAAAGGAA AGTGGTAAAG ATGGAATT TCTGTCAG 13650  
 ATTTTAGTCT CCCAACACAG AGCTCAGATG TGAGAAATG TGCAAGACTG TGAGACAGAA CAATAGACAA 13720  
 AAACCGAAGC CCTATCTCTC AGAAACCTGTG TTAAATGTTG TATGTCAGC AGCTGATGGA AAAAGAGGTG 13790  
 50 TGTGTGTAAT TTTTTTCTC GAGAAAAGTC ACTGGAAAGC AATAAGTTG TGCTTTACAG CATATACCA 13860  
 AGCAGATTCT AGGTAGAAGA GGAGACACAT GCAAAACACA CCAGCAGAC AAATAAAACA AARGACTCAA 13930  
 AGGGAAAGGGG GGTGAACGCTT CCTCTGGTTG TGTTGGGGAGGACACAGA GGGAGGGGA TGAAGGAGGT 14000  
 GAGGAACCGG GCATTGCTT CACTGCAGAG AAACCTCAGCT TGCCCTGAGG ACAGTGAAGA TGCCATCTC 14070  
 CTGGAGCTGT TGTGACGCTG ATTATTTAA TGCGGGCTCTG GAGGGCTCTG ACATTCATCC TCTCACTTT 14140  
 55 TTCTCTAAC CACCTGGAGG TGAGGAGGGG AAAGGCTCCA GGGGGAGCAGC CGGGCTCTGG CACCCAGCTG 14210  
 GCAAAAGGGCA TGCACTGATT CAGGGCTGGC TGCTGCTCC GGGGGCTCTG TGCTGCCCCAG GACCCCAACAC 14280  
 AAGTCAGACG CATAGGCTCA GGGTGGAGGG GAGGCCAAGG TGCTGTTGGG GATGGCTGTG AAAGAAGAAA 14350  
 TGACACTCTG GGGGAAGGCTT CCTCTGGTTG TGTTGGGGAGGACACAGA GGGAGGGGA TGAAGGAGGT 14420  
 AGACCCCATC CTCAAAAGAA CGCAGCTGGAG AACTCAGCT TGCCCTGAGG ACAGTGAAGA TGCCATCTC 14490  
 TTTCTGGGG TTGCAAGAG CCAGCTCATG TTGAGGCAA CTGGAGGAA CTGGCTCTG AAGCAGCTTG 14560  
 60 TTTCATGGA AGTCTCTACA ATGCTCTGTG TCTCTCCAGT AATTCACCTT CTGAAGTGTAC CAGACATTAT 14630  
 CACGGTCTT ATTTCACATT TCCAGTGTG CAGGGAGGGG GAGCTGGCAC AGCAAGTCAAC GAACTCTGCC 14700  
 AAATACAGGG CTAAGGAGAT ATTATGACAT CAAACAACTT CTCTGGCATT AAACATTTC CAAAGAATT 14770  
 TTGAAGGATG TTAATGGCA AAAACGTTT ATTCAATG TGCACTGGTCA AAGACTGGAT GTAAAAGAAC 14840  
 65 ACACCCCGAGG AGGCTGCGCT GAATGCTGAG TGCTGTTCTG TTGGACATG GACATACATG GGCACTGGT 14910  
 GGTGGTGGAGG CCTCTGGAGG CATCGCTGGG ATGCCCTCATC CCTGGCCCTC TGAGGACACC ATGTTGCCA 14980  
 CGTGCACCA CTGGGGCCCT GTTTAGTGTG TGCCACCTGG CTCTTCCATC CCTGAGATTC AAACACAGTG 15050  
 AGATTCCTCA CGGGCAACTC AGTGGTCTCC CACCTTCAAC CGGGGCTTC AGGCTGGGGAG GAAACATGGT CGCAAAACTCT 15120  
 GCGGGGGAGC CAGGGCTCCA CAGTTTATA TGTTTTTG TGCTGAGTTAT GTGCAAGACT CATCAGGGCA 15190  
 70 GATGATGAGT GCACAAACAC GGCGCTGCGA GTGTTGGATA CACTCAACAT CACTAGCCAG GTCTGGTGG 15260  
 AGTTTGGTCA TGCAAGACTG GGATGGCAGT TGCACTGGAG AGTCCATGGA GTGAGCACC AGGCCCCCTCG 15330  
 GGGTGCAGCG CAGGGCCCG AGCAGGACAAG GAAAGGGAGG GAAAGGAGGA GGCTCTTGG AGCAAGCTTT 15400  
 GCAGGAGGGG GCTGGGTGTG GGGCAGGCACT CGTGTCTGA CATTCCCCC TGTGTCTGAG CTATGCCCG 15470  
 ACCTCCATCA GAGCCAGTCT CACCTTCAAC CGGGGCTTC AGGCTGGGGAG GAAACATGGT CGCAAAACTCT 15540  
 75 TTGGGGCTT GCGGGCTGAAG TGTCACAGCC TGTTCTGGA TTGCGAGGT AGCAGGCTGA TGGTCAGAC 15610  
 AGAGTTCAAG GTTCAAGGAGG TGTTGCGCA AGTATGTTG TGTTGTTG TGCAAGGCTG 15680  
 ATGGTGAATG GCTGCACGTA AGAGTGCACA TGACGGATA TACACGTGAG CACATACATG TGTCATGTTG 15750  
 TGTCATGAA GGCACTGGCAAG GTGTTGCAAG GGCACAAAGTG TGTCACATG CGAATGCACA 15820

	CCTGACATGC	ATGTGTGTT	GTGCACAGTC	GTGTGGGCAT	TCACGTGAGG	TGCATGGTG	TGGGTGTGCA	15890
	GTGTGAGTAG	CATGTGTGCA	CATAACATGT	ATTGAGGGGT	CCTCGTGTTC	ACCCCGCTAG	GTCTCAGCA	15960
5	CCACTGCCAC	TCTTACAGG	ATGAGACGGG	GTCCCAGGC	TTGGTGGGCT	GAGGCTCTGA	AGTGCAGCC	16030
	CTGAGGGCAT	1GTCCCATCT	GGGCATCCGC	GTCCACTCCC	TCTCTGTGG	GCTTCTGTGT	CACTCCCC	16100
	TCTCTGTGG	GCATTACAT	CCACTCCACT	CCCTCTCTCC	TGTGGGCATC	CCGTCACACT	CCCCCTCTCT	16170
	GTGGGCATCT	GGGTCCACCT	CCCTCTCTG	TGGGCATTG	CGTCCACTCC	CTCTCTGT	TCCTTCTGT	16240
10	CTTGGCCAGG	CCTGGGGGC	AGGCAGATGA	CACAGACTCT	TGACTCGCC	AGGGTGGTTC	GCAGCTGCCG	16310
	GGTGAGGGCC	AGGGGGGATT	TCACTGGGAA	GAGGGATAGT	TTCTTCCCTC	TTCTTGTGTC	16380	
	CATCTGAATG	ATGATAAAG	AAAAAAATAA	AAACCTTTAA	CCACAGAGAG	TTTCTACCG	TCTCTACTC	16450
	TTTCTGGG	ACTCTAGGT	AAACGGCTCC	AGACGGTGT	CACCAACATC	TACAAGATCC	TCTGCTGCA	16520
	GGCGCTACAGG	TGAGGCCGCC	CCAAAGGGGT	CAGGGCCAGC	CTCCGGGCT	CTGCTCACCT	16590	
15	CTGACCGGGG	GGTCACCTT	GGAACTCTG	GTTTTAGGG	CGTCCACTCC	CTCTCTGT	TCAGTGTGTC	16660
	TGCTGCCGT	GCACAGTCT	GTTCGGTGTG	CTCTGTGAA	AGCACCTGTT	CTCCATCTCT	GGTAGTGTG	16730
	AGGAGCCGGT	GTGGCCCGAC	TGTGGCCAC	TGTGGCCAC	CACTGGCCCT	GGGACGTCAT	GGAGGCCATC	16800
20	CCAGGGCAGC	AGGGCATGG	GGTAAAGAGA	TGTTTATGGG	GAGTCTTACG	AGAGGAGGCT	GGGAAGGTTG	16870
	CTGAACTGTA	GATGGGAGAT	CAGATGCCG	GAGGATTGGG	GGTCTCAGCA	AAGAGGGCCG	AGGTGGTGC	16940
	AGGTGAGGGT	CCTGGGCCCC	ACCCCGGGGA	AGGTGACCA	GAGGTGGGAC	TCCCCACACA	GCCCGGGCCAG	17010
25	CACCTGTGCT	CTGGGATGG	CTGTCCTCT	GGAACTTTC	CTGTCCTGG	TGTTCAAGGG	GTGCCCCCTGC	17080
	CAAGAAATCGA	CAACTTTATC	ACAGAGGGAA	GGGCAATCT	GTGGAGGCCA	CAGGGCCAGC	TTCTGCCG	17150
	AGTCAGGGCA	GGTGGTGGCA	CAAGGCTCGG	GGCTGTACCA	AAGGGCAGTC	GGGACCCACA	GGGCCCCGGC	17220
30	TCCACCTCAA	CGACCCCTCC	GAGCCACTG	GAGCTGAATG	CCAGGAGGCC	GAAGGCCCTC	CCCCATGAGG	17290
	GCTGAGAAGG	AGTGTGAGCA	TTTGTGTAC	CCAGGGCCGA	GGCTGCGCA	ATTACCGTGC	ACACTTGATG	17360
	TGAAATGAGG	TCGTCGCTCA	TCGTFGAAC	CCAGCAAGGG	CTCAGGGGAG	AGTTTCTCAT	TACAAGGTG	17430
35	TACCATGAAA	ATGTTTTTA	ACCCGAGTGC	TTGCGCCCTC	ATGCTCTGG	AGGGAGGGCA	GAGCCCACAGC	17500
	TGCAATGTTAC	CGCCTTTGCA	CCAGCTCCAG	AGGCTTGGGA	CCAGGCTGTC	TCAGTTCAG	GSTGCGTCCG	17570
	GCTCAAGACCC	CCCTCTCTC	TGCGCTCTC	CTCTGCTCTA	ATCTTCCCTC	CCTTGCATC	TCCCTGACGC	17640
40	GTGCGCTGGC	CCTGTCGAA	GCTGCTGAC	TCTTCTTCCC	AAACCTTGGG	GGTGTGCTGG	ATACAGGTG	17710
	CACTGAGGAC	TGGAGGTGTC	TGACACTGTG	TGTTGACCCCA	GGGTCAGGCT	GGCGTGTG	GGGCTCTCTT	17780
	GGGGCATGTG	GAGGCTCAGG	GAGTTTCCC	AGGTGAAAC	TCTTGGGAA	CTCCCAAGGG	CATGTGACCT	17850
45	GGCACCTGCT	CCTCCCATAT	TCACTCTCAGT	TCTGTCCTC	TTTCCCAC	AGGGTCTCTA	GCTCCAGAGA	17920
	GCTCCGTAG	AGGGCCTGGG	CTCAGGGCAG	GGCGGCTGAG	TTTCCCAC	CATGTGGGGA	CCCTTGGGTA	17990
	GTGCGCTGAT	TGGGGTAGCC	TGAGGAGGCC	GAGATGCGAT	GGGCAACGGG	CCGTTTCAA	ACACAGACTC	18060
50	AGGCACGTGG	AGGGCCAGG	AACTCCCTC	CTCTGAGGCA	GGAGTGGGAG	AACGGAGAGC	TGGGCCCCGA	18130
	TTTCAGGGCA	GCCAGGCTGC	AGTGGGCGAG	GCTGTGGTG	TCCACGTG	GCTGGGGCGC	GGGTCTGATT	18200
	CAAATCCGCT	GGGGCTCGG	CTTCCCTGGG	CGTGTGGG	GGGGCTCCAC	ACGGGCTGG	GGTGGACGCC	18270
55	CCGACCTCTA	CGAGGTGCT	ATTCTCTCT	TTGGAAGAGA	GGCCCCTCACC	CATGCTAGT	TTTCCCCTC	18340
	TGGGTAGGA	CGCTGGCGGT	GTGGAACCC	CGGGACCTTA	GGCTTATTAA	TTTGTGTTAA	AAACTCTGG	18410
	GCCCTGGCTC	CGTGTGTTCT	AAATGGGGAA	AGAACATCCC	ACCTCAGCAG	AGTTACTGAG	AGGCTGAAC	18480
60	CGGGGTGCTG	CGTGTGACTGG	AAATGGGGAA	TTTGGGAGGAT	ATGGAGGCTG	ACAGGGCTGG	CGCGGTGGCT	18830
	TACATGGGG	GCTCAGGCAG	TGGGTGAGAT	GAGGTACACG	GGGGGCTAG	GCAGTGGGTG	AGGCCAGGTA	18620
	CATGGGGGGC	TCAGGCAGTC	GGTGGAGATG	GGTACACGGG	GGGCTCAGGG	AGAGGGTCAG	ACAGGGTACA	18690
	CGGGGGCTCT	GATCACACGC	ACATATGAGC	ACATGTCAC	ATGTCGTTG	TATGGTAGC	CAGGCTGTG	18760
	CACACCTGCC	CCAAAGTCCC	AGGAAGCTGA	GAGGCAAAG	ATGGAGGCTG	ACAGGGCTGG	GGGCTGGGCT	18830
65	CACACCTGTA	GTCCACGGAC	TTTGGGAGG	CGGAGGAGG	GGATACCTCTG	AGCCCAGGAG	TTTAAGACCA	18900
	GCCTGAGCAA	CATAGTAGAA	CCCCATCTCT	ATGAAAGATA	AAACACAAA	TTAGCTGAC	ATGGTGGTGT	18970
	GCGCCTGTAG	TTCCAATACT	TGGGAGGCTG	AAAGTGGGAGG	ATCACTTGAG	CCCAGGAGGT	GGAAAGCTGCA	19040
	GTGAGGTGAG	ATTGACACCA	TGTCATGCA	GGTACACGGG	GGGCTCAGGG	AGAGGGTCAG	ACAGGGTACA	19110
70	GAAGACTGAC	AAATGCACTG	TCTTGGGAA	AAACATTAG	TAGGAACCTTA	ACCTTACAC	AGAACCCAA	19180
	TCGGTGTCTC	GGTGTGACTG	AGATGAGATG	ATGGGTCTC	ACACCATCAC	CCCAGACCA	GGGTTTATGCC	19250
	ACACACGGGG	CGGGTGGCTC	AGAAGGGATG	CGCAGGACGT	TGATATACGA	TGACATCAAG	GTGTCG	19320
75	GAAGGGCAGG	ATTATGATA	ATGACCTGCT	GGTACACAGA	GAACAACTGGA	TAACATGGAA	ACCTTACAGG	19390
	CCTCCCGGA	ACAGGGGCTA	ATCAGAAGCC	AGCATGGGG	GCTGGCATCC	AGGATGGAGC	TGCTTCAGCC	19460
	TCCACATCGG	TGTTCATACA	GATGTGTCAC	GAAGACCGC	TGACATCTG	CACACACAGA	CRGGCAGCTA	19530
80	CTGGCACACA	CAAGCACACA	CACAGACATG	CTGTCATGCA	TCCGTGTG	TGCACTCTG	CCCATGAGGA	19600
	AACCCATGCA	TGTGCATTCA	TGCAACGACA	CAGGCACCGG	TGGGCCCCATG	CCCACACCCA	CGAGCACCCG	19670
	CTGATTAGGA	GGCCCTTCTC	CTGACGCTGT	CGGCCATCT	CTCAGGTTG	ACGGCATGTT	GCTGCAGCT	19740
	CCATTCTCATC	AGCAAGTTG	GAAGAACCCC	ACATTCTTC	TGGGGCTCAT	CTCTGACACG	GCCTCCCTCT	19810
85	GCTACTCCAT	CCTGAAAGCC	AAAGACCGAG	GTATGTCAG	GTGCTGGCC	TCAGTGGCAG	CAGTGGCTGC	19880
	CTGCTGGTGT	TAGTGTGTC	GGAGACTGAG	TGATATGCTG	CTTGAAGGAT	TCTTACCCCT	TTTCGATCA	19950
	GGAAAGTGGT	TAACCCAAAC	ACTGTCAGGG	TCTGCTGCC	GGCCCTCTG	GGGGTGGAGCA	GAGCAGCTGA	20020
90	TGGAAGGGAC	AGGAGCTGTC	TGGGAGCTGC	CATCTTCCC	ACCTTGTCT	GCCTGGGAA	GGGCTGGGGG	20090
	GCCTGGCTC	TCTCTGTTGC	CCCCATGTTG	GATTGGGGGG	GCTGCCCC	TCTGTTG	CCTGTTG	20160
	GATTGGGGTG	TCTCCCGTC	ATGGCACTTA	GGGGCCCTTG	GCACACCGG	GGCAAGGGCT	TAGGAGGAGG	20230
95	CCAGGCCAG	GCTACCCAC	CCCCCTCAGG	AGCAGAGGCC	GCGTATCAC	ACGCAAGAGC	CCCCGGCCCGT	20300
	CCTCTGCTTC	CCAGTCACCC	TCTCTGCC	CTGGACACTT	TGTCCAGCAT	CAGGGAGGTT	TCTGATCCGT	20370
	TTGAAATTCA	AGCCATGTCG	AACTTGGGGT	CCTGAGCTTA	ACAGCTCTA	CTTCTGTT	TTTGTGTT	20440
	GTGAAATTTC	CACCTGGAGA	AGCCGAAGAA	AACTATTCCT	TGTCAGCTCC	TGCGGTGCTT	GGGTCGGGAC	20510
	AGCCAGAGAT	GGAGCCACCC	CGCAGACGGT	CGGTGTGGG	CAGCTTCCG	GTGTCCTCT	GGAGGGAGGC	20580
100	GGGGCTGGCC	CTGTGACTCC	TCAGCCTCTG	TTTCCCCCA	GGGATGTCGG	TGGGGGCCAA	GGGGCCGCC	20650
	GGCCCTCTGC	CTCTCGAGGC	CGTGCAGTGG	CTGTGCCCCA	AAAGCATTCCT	GCTCAAGCTG	ACTCGACACC	20720
	GTGTCACCTA	CGTGCACACT	CTGGGGTCAC	TCAGGACAGG	CAAGTGTGGG	TGGAGGGCAG	TGGGGCCCC	20790
	ACCTCCCCAG	GGGTCACTCT	TGAACGGCT	GTGTTGGGG	AGCAGCTCTA	GATGCTGCTG	AAAGTCAGAC	20860
	GCCCCGGGGC	CTGACCCCTG	GGGGCTGGAG	CCACGGCTGG	AGCCCTATG	GATTAAACGG	TGGTGTGCCC	20930
105	AGGCCACGGA	GGCTGGCAGG	GTCCCCAATC	TCTTGAACCC	CTGTCCTCTA	TCTCAGGGGC	GATGGCTCCC	21000
	CACGCTTGGG	AGCCTCTGTA	CCCCCTGACCT	GTGTCTCTC	ACAGCCTCTT	CCCTGGCTGC	TGCCCCCTGAGC	21070
	TCTCTGGGTC	CTGAGCAAGT	TCTCTCCCG	CCCCGGCCGT	CCAGCGTCAC	TGGGCTGCT	GTCTGCTGCG	21140
	CCCCGGTGGAG	GGGTGCTGT	CCCTTCACTG	AGGTCCAC	CAGCCAGGGC	CACGAGGTG	AGGGCCCTGCC	21210

5	TGCCCCGCCA CCCACRCGTC CTAGGAGGGT TGGAGGATGC CACCTCTGGC CTCTTCTGGA ACGGAGTCG 21280 ATTTGGCCCG CGCAGCCCCAG ACGCAGCTGA GTCTGGAAGCT CCCGGGGACG ACGCTGACTG CCCTGGAGGC 21350 CGCAGCAAC CGGGCACTGC CCTCAGACTT CAAGACCACAT CTGGACTGAT GGCCACCCGG CCACAGCCAG 21420 GCCGAGAGCA GACACCAGCA GCCCTGTCA GCGGGCTCT ACCTCCCAGG GAGGGAGGGG CGGCCACAC 21490 CCAGGCCCCG ACCCGTGGGA GTCTGAGGGC TGAGTGAAGT TTTGGCCCGAG GCCTGCAGTG CGGGCTGAAG 21560 GCTGAGTCG CGGCTGAGGC CTGAGCAGAGT GTCCAGCAGCA GGGCTGAGTG TCCAGCACAC CTGGCGTCTT 21630 CACTTCCCCA CAGGCTGGCG CTGGCTCCA CCCCAGGGC AGCTTTCTT CACCAGGAGC CGGGCTCCA 21700 CTCCCCACAT AGGAATAGTC CATCCCGAGA TTCCAGGATTG TTCAAGCCCTC GGCTTGCCCT 21770 CCACCCACAC CATCCAGGTG AGAACCTGA GAAGGACCTC GGGAGCTCTG GGAATTGGA GTGACCAAAG 21840
10	GTGTGCCCTG TACACAGGGC AGGACCTGC ACCTGGATGG GGGTCCCTGT GGGTCAAATT GGGGGAGGT 21910 GCTGTGGGG TAAATACTG AATATATGAG TTTTCAGTG TTGAGTAAAGT TCTCATGTTT GAATCTTAAT 21980 GTGCACTGCA TAGACACAC TGTATGCAAT TACAGAACGC TGTAGTGAAG CGGGGTGGTG GTCACTGCGG 22050 GCCCATGGCC TGGCTGTGCA TTACGGAGG TCTATGAGTG AATGGGGTTG TGGTCACTGCG GGGCCCATGG 22120 CCTGGGAGGT CCTGGGAGGT TTCTGATGCT GTAGGGCAGG AGGGGAAGGA GGGTAGGGGA TAGACAGTGG 22190 GAGCCCCCAC CGTGGAGAGC ATAACAGTAA CCTCAGGCCA GAAGGAGCAG AGGGATGCTG GGGGCCACAG 22260 TTGGCGGGC GGGATGATGG AGGGCTGGC CAGGGTGGCA GGGATGATGG GGGCCCCAGC TGGGGTGGCA 22330 GGGGTGTGGG GGGGGCTGGC TCTGGGGCTGG GGGGAAGATG GGGAGGCGC GCTGGGGCCC CTCCCTCCCT 22400 GCCCTCCACC TGACGGCGT GATCCGGATG TGCTTCCCTG TGTCACATCC TCTGGGGCAT CAGCTTCTAT 22470 GGAGGTGGGG GGCAGGGGCA TGACACCAC TGTATAAAA TCCAGGATTG CTCCCTCTGA ACAGCCCCAAC 22540 TCAGGGTAA AGTCACATTC CGCCCTGTGCA CATTCTGTTA AGAGTAGACG AGGATTCGAGA TCTCTGAAGG 22610 TGGGGTAGGG TGCGGAGTGG GAGGGTGGG ACACAGGGG CTTCAAGGGT GGGCTGGTG TGCTCTCTCA 22680 TCTCTTATC ATCTCCAGT CTATCTC ATCTCTTAT CATCTCCAG TCTCATCTG TCTCTCTTA 22750 TCTCTCCAGT CTATCTGTC TCTCTTACG ATCTCCAGT CTATCTCTG ATCTCTTAT CATCTAGTCT 22820 25 CATCCAGACT TACCTCCCG GGGGGTGGC AGGCTGGCAG TGGAGCTGGAA CATACTCTC TCTCTAGGCA 22890 GAAGGAACGT GAAGGATTGC AGAGAACRGC AGGGGGCGCT CAGAGGGAGC CAGTCTGGG GTGAAGAAC 22960 AGGCCCTCTC CAGAATTGGT CTGGGGCCAG ACAGAACCCAG GGGGGGGGGC TGAGTGGCTC CAGAGCTTC 23030 CAGCAGGTGG CGTGGGGGGC TTATGGTAT GGGGGGGCTC TACTGAGTGC ACCTTGGACA GGGCTCTGG 23100 TTTGAGTGC GCGGGACGT CCTGGTGTG GGGGGGGG CTTATGGCCA CTGGATATGG CGTCATTAT 23170 TGCTGCTGCT TCAAGAAATG TCTGAGTGC CGAGCTTAAT GTGATGTTG GGGCCAAAGTC CACAGACTGT 23240 30 GTCGTAATG CACTCTGGT CCTGGGAGCC CCGTATAGGA GCTGTGGAGGA AGGAGGGGCT TTGGGAGGC 23310 GGCCTGGGG CGCCTTGGC CTGCAAATCG GAAGGGAGG GCCCCGGGGC CGTGTGGGGG ACCACCTCAA 23380 GTGAGGGTT GGACAGAACAG GGGGGGGGAC TTCCCAGGAG CAGAGGGGGC TGTCAGGGCA CACCTGGGGT 23450 TGAATCACAG ACCAACAGGT CAGGGCATTC TTCACTATC CATCTCTAC AAGAGCTTCA ATTCTCTTT 23520 35 CTCCGGGTG TTTTGTGA AATTTTACTC AGGATTACTT ATATTTTG CTAAAGTATT AGACCTTAA 23590 AAAAGTATT TCGTTTGATA TGCTTAACTC CACTAACGC CACTTTAT TGTCTGTTT TATTTTATT 23660 TATTATTATT ATTAGAGATG GTGCTACTC TGTCAACCGG CTTGTTAGTG CAGTGGCACA GTCATGGCTC 23730 GCTGTAGCCG CAAACCCCCA GGCTCAATG ATCCCTCCGGC CTCAGCTTCC CAGAGTGTG GGATTAACAGG 23800 TGTGAGCCAC TGGCACTT AAAAACACT ATGTAAGGCT AGGTCCAGTG GCTTCCACAC 23870 40 TGTCATCCC AGTAGTTGG GAAGCCAGGG CAGAGGATTG TGCTGGGGC AGGAGTTGA GACCAGCATG 23940 GGTAACATAG GGAGACCCCA TCTCTACAAA AAATGCAAAAG ATTTATGGG CGGTGGGGTC CAGCATCTG 24010 AGTCCCCAGCT GCTCGGGAGG CTGAGTGGGAGG GGATCGGGT AGCCCCGGAG GTCATGGCTG CAGTGGAGCTG 24080 TGATTGTACCC ATGCCACTCC AGGCTGGGCA ACAGAGTGAACCCCTGTC AAAAAAAA AAAAAAAAG 24150 AAGGAGAAGG AGAAAGAGAAG AAGAAGGAGG AAGGAAAGAG AAGAAGAAGG AAGAAGGAAG AGAAAGGGAG 24220 45 AAGGAGGCCT GCTAGGTGCT AGGTAGACTG TCAAAATCTCA GAGCAAAATG AAAAAAACAAAGTTTAAAG 24290 GGAAAAGAAA ACCCCAGCTC TTGGACTTC CTTAGGGCTG AACTCTCATC CAGAGCTT CTTCCACAG 24360 ACAAGGGTGT ATGGAGGGAG TGAGTTCAAA GCAGAAAGGG AGGAGAAGGC GGCAGGGGTG GAGGCTGTGG 24430 GTGACACCAG CGAGGACCCC TGAAAGGGAGG TGGTTGTTT CCTGCTCTAG CCCCACGCTC CTGGGGTTC 24500 TGACCTCTGCT GTAAACCGTCC ATGTTGGTC CAGGTGGCCA CCTGGGGAGG ATGCTGTGCA GGGGGCTTC 24570 50 CAAACTTGG TGGTTTCAG AAGCCCCAGG CACTTGTGGC AGGCACAATT ACAGCCCCCTC CCCAAAGATG 24640 CCCACCTCTC TCTCTGGAA CCTGTGAATG TGTCAACCCCG AAGGAGAGG CTGGTGAAGG CTGCAGGTGG 24710 AATCACGGCT CGCAGTCAGC CGATCTTAAG CTACCTCTGG ATTATCTGGT GGGGGCTGATA TGCCACACAG 24780 GGTCCCTAGA AGTGAGAGG GGAGGCAGGG GAGAGTCAGA GAGGGGAGGT GAGAAGGACCC ACTGGCCACT 24850 GCTGGCTTGG AGATGGAGG GGGGGGGGGG AGGCAAGGGAGA TGGGGGCAGC CGCTCTATGC TGAAAAGGCA 24920 AGCAATCTCTC CCCGGCTCTG AGGGGCACAGC GCCCCCTGCCA CGCCCTGCTG TCAAGGGAGT GGGACCTGTT 24990 55 TCAGCTTCTC GGCCCTCCAGA GCTGTAAGAT GATGGGTTTG TGTTCAGGCA CTAAGCTGCA GTGATTCTGTC 25060 ACAGCAGCAA ATGGAATAGC AGTACAGGGAA ATGAAATACA GGGAGCTTC TCAAGGTGAC TCTCAGCCCCA 25130 CCCCCTGGG 25138

60 **Beispiel 5**

Der Vergleich der oben beschriebenen genomischen hTC-Sequenz mit der Sequenz der hTC-cDNA (Fig. 6; entsprechend SEQ ID NO 2) ermöglichte die Aufklärung der Exon-Intron-Struktur des hTC-Gens. Die genomische Organisation des hTC-Gens ist 65 in Fig. 7 schematisch dargestellt. Die kodierende Region des hTC-Gens setzt sich aus

16 Exons zusammen, die in ihrer Größe zwischen 62 bp und 1354 bp variieren (s. Tabelle 1). Exon 1 enthält das Translationsstartcodon ATG. Das Translationsstopcodon TGA sowie der 3' untranslatierte Bereich liegen auf Exon 16 (Fig. 8). Ein mögliches Polyadenylierungssignal (AATAAA) wurde weder in Exon 16 noch in den 3195 bp der folgenden 3'-flankierenden Region gefunden. Basierend auf der Konsensussequenz

	5'-Exon				Intron				3'-Exon							
	Prä-mRNA	A/C	A	G		G	T	A/G	A	...	N	C	A	G		G
10	Häufigk. (%)	70	60	80	100	100	95	70			80	100	100	60		

wurden die Exon-Intron-Übergänge bestimmt und in Tabelle 1 aufgeführt. Mit Ausnahme der 5'-Splice-Stelle zwischen Exon 15 und Intron 15 stimmen alle Exon-Intron-Übergänge mit der publizierten (Shapiro und Senapathy, 1987) Splice-Konsensussequenz überein. Die Größe der Introns liegt zwischen 104 bp und 8616 bp. Da Intron 6 nur zum Teil isoliert wurde, kann die exakte Länge des hTC-Gens nicht bestimmt werden. Basierend auf der von Intron 6 erhaltenen Teilsequenz von ~4660 bp beträgt die minimale Größe des hTERT Gens 37 kb.

Die Introns 1-5 sowie der 5'-Bereich des Introns 6 sind in Contig 1 enthalten:

Intron 1: bp 11493-11596 (SEQ ID NO 4);  
Intron 2: bp 12951-21566 (SEQ ID NO 5);  
Intron 3: bp 21763-23851 (SEQ ID NO 6);  
5 Intron 4: bp 24033-24719 (SEQ ID NO 7);  
Intron 5: bp 24900-25393 (SEQ ID NO 8);  
5'-Bereich von Intron 6: bp 25550-26414 (SEQ ID NO 9).

Der 3'-Bereich des Introns 6 sowie die Introns 7-15 sind in Contig 2 an folgenden

10 Positionen lokalisiert:

3'-Bereich von Intron 6: bp 1-3782 (SEQ ID NO 10);  
Intron 7: bp 3879-4858 (SEQ ID NO 11);  
Intron 8: bp 4945-7429 (SEQ ID NO 12);  
Intron 9: bp 7544-9527 (SEQ ID NO 13);  
15 Intron 10: bp 9600-11470 (SEQ ID NO 14);  
Intron 11: bp 11660-15460 (SEQ ID NO 15);  
Intron 12: bp 15588-16467 (SEQ ID NO 16);  
Intron 13: bp 16530-19715 (SEQ ID NO 17);  
Intron 14: 19841-20621 (SEQ ID NO 18);  
20 Intron 15: 20760-21295 (SEQ ID NO 19).

Der 3'-nichttranskribierte Bereich befindet sich ebenfalls im Contig 2 an Position 21960-25138 (SEQ ID NO 20).

25 Die genannten Introns haben im einzelnen folgende Sequenzen:

## Intron 1 (SEQ ID NO 4)

GTGGGCCTCCCGGGTCCGGCTGGGCTGGGTTGAGGGCGCCGGGGAAACCAGGACATGCGGAGAGCAGCGCAGG  
CGACTCAGGGCGTTCCCCCGCAG

5

## Intron 2 (SEQ ID NO 5)

GTGAGGAGGTGGTGGCGTCGAGGGCCCAGAGCTGAATGCAGTAGGGCTCAGAAAAGGGGCAGGCAGAGCC  
CTGGTCTCTGTCTCCATCGTCAGTGGCACAGTGGCTTCGCTCAGGACGTGAGTGGACACGGTGATCTCTGCC

10

TCTGCTCTCCCTCTGTCCAGTTGCATAAACTTACGAGGTTCACCTCACGTTGATGGACACGCGGTTCCAGGCC  
CGAGGCCAGAGCAGTGAACAGAGGAGCTGGCGCGCAGTGGAGCCGGTTGCCGGCAATGGGAGAAGTGTCTGAAG  
CACAGCCTCTGGCGAGGGTGCCTGCAGGTTACCTATAATCCTCTGCCAATTCAGGGTGGGAATGAGAGGTGGGA

15

CGAGAACCCCTCTCCGGGGTGGAGGTAAAGGTTTGAGGTGACGGTGCACGTGGTCAGCCAATATGAGGTTGTTA  
AGATTAATTGTGTGTTGACGGCCAGGTGCGGTGGCTACGCCGTAATCCAGCACTTGGGAAGCTGAGGCAGGTGGAAG  
TCACCTGAGGTGAGGAGTTGAGACCAGCTGACCAACATGGTAAACCCATCTGTACTAAAAATACAAAATTAGCTG

20

GGCATGGTGGTGTGCGCTGAATCCAGCTACTTGGGAGGCTGAGGCAGGAGAACTCTGAAACCCAGGAGGCGAGG  
TGCAGTGAGCTGAGATTGTGCCATTGACTCCAGCCTGGCGACAAGAGTGAACACTCTGTCTTAAAAAAAGTGT  
CGTTGATTGTGCCAGGACAGGTAGAGGGAGGAGATAAGACTGTTCTCAGCACAGATCCTGGTCCATCTTAGGTAT  
GAAGAGGGCACATGGGAGCAGAGGACAGCAGATGGCTCCACCTGCTGAGGAAGGGACAGTGTGTTGGGTGTTAGGG  
ATGGTGTGCTGGGCCCTGCCGTGCCCCACCCCTGTTTCTGATTGATGTTGAGGAACCTCCGCTCAGGCCCTT

25

TGGCTCCAGTGTCCCAGGCCCTACCGTGGCAGCTAGAAGAAGTCCGATTTCACCCCTCCCCACAAACTCCAAAGAC  
ATGTAAGACTCCGGCCATGCAGACAAGGAGGGTACCTCTGGGCTCTTTTTCTTTTTCTTTATGGTGGC  
AAAAGTCATATAACATGAGATTGGCATTCTAACACCGTTCTGTGTACAGTGCAGAATTGCTAACCGGGTGT  
CAGCAGGTTGCTGAAATGCTGCGCTTGTGACTGGAAGTCCCTACCCATGAAACGGCAGCTGCTCACACCTGCTGC  
GGCTCAGGGGACACGCCAGTCAGATAAGCTCATGCAACCCAGTTGCTTGTGCTCCAGCTTCTCGTGAG

30

GAGAGTTGAGTTCTGTGATCAGGACTCTGCTGTATTGCTGTTCTGACTTCAGATGAGGTACAATCTGCCCTGG  
CTTATGCAAGGGAGTGAGGCGTGGTCCCCGGGTGCTCCCTGTCACGTGCAAGGGTAGTGAGGCGTTGCCCTAGGTG  
GTCACGTGAGGGTAGTGAGGCGCGGCCCCGGGTGTCCTGTCCCCTGTCAGCGTGTGATTGAGGTGAGGCGTGG  
CCCTGTCACGTGAGGGTAGTGAGGCGCCATCCCCGGGTGTCCTGTCACTGTGAGGGTAGTGAGGCGTGGTCCCCGG  
GTGTCCTGTCCCCGTGCAAGGGTAGTGAGGCGACTGTCCCCGGGTGTCCTGTCACTGTGCAAGGGTAGTGAGGCG  
CCGGGTGTCCTCTCAGGTGAGGGTAGTGAGGCGCGGCCCCAGGGTAGTGTCACGTGAGGGTAGTGAGGCG

35

GTCCTGGGTGTCCTCCAGGTATAGGGTAGTGAGGCGACTGTCCCCGGGTGTCCTGTCACTGTGAGGGTAGTGAG  
CCGGGCCCCGGGTGTCCTCTCAGGTGCAAGGGTAGTGAGGCGCTGTCCCCGGGTGTCCTGTCACTGTGAGGGTAG  
GAGGTCTGCCCCAGGTGCTTGGCTTGCTACCTGAGCTTGTGCTCTGCAATGTTGCTCTTATAGCCACAGCT  
GCCCGGTTGCCATTGCTGGTAGATGGTGCAAGGGCAGTGTGTCAGGCTTGTGCTCCCCAGGCTATCTTGT  
TCTGGTCACTCTCCGTTCCATTGGTAGTGAGGGACACGGGACTGCAGGCTCTGCCCTCCGGTGCAGGCACTGCA  
CCACAGCTCAGGTCCGCTCTGTTGGGCTGGCTTGCTCACAGTGGCCACATGCACTGCAATACTCC

40

TCTCCCAGCTGTCTCATGCCAGGGTAGTGAGGCGACTGTGGGCTGCTGCTGCCACGTGCTGGAGACATCCCAGAA  
AGGGTTCTGTGCCCTGAAAGGAAAGCAAGTCACCCAGCCCCCTCACTGTCTGTTCTCCAAAGCTGCCCTCTGC  
TTGGCCCCCTGGGTGGTGCAAGCTTGTCACCTTATTCTGGCACCTGCCGCTATTGCTTAGGCTGGCTG  
CCAGTCGCCCTCACATGGATTGACGTCCAGCCACAGGTTGGAGTGCTCTGTCTGCTGAGACCCACGTG

GAGGGCCGGTCTCCGCCAGCCTCGTCAAGACTCCCTTGGGCTTAGTTGAATTCACTGATTACCTCTGACG  
 TTTCTATCTCCTTCATTGTATGCTTTCTGGTTATTCTTCATTCCCTTCTAGCTCTTAGTTAGTCATGCCCTTC  
 CCTCTAAGTGCCTACCTGCACCCCTGTTGATGTGAAGTAATCTAACATCAGCCACTTCAGTGTCTTAA  
 ATACTCAAAGTGTAAACTCTTTAAGTATTCTTATTCTGTGATTTTCTTGACGCTGTGTTGACGTGA  
 5 AATCATTTGATATCAGTGACTTTAAGTATTCTTAGTATTCTGTGATTCTTGAGCTAGTGTGAGTTATTGAAACT  
 GTTTATGTTCAAGATATGTAGAGTATCAAGATACGTAGAGTATTAAAGTTATCATTATTGATTCTAACTCAGT  
 TGTGAGTGGCTGTATAATACCAATTGAGTTGCGGAGCCTGCTTGATCTAGTGTGATGGTTCCAG  
 AACTGTCCATTGAAATTGACATCCTGCAATAGGGCATGCACTATCCAGCTTATAAGGTCCAGTGCA  
 AAGCTTCTGCTCCTCTAGATGCATGAAATTCCAAGAAGGAGGCCATAGTCCCTACCTGGGGATGGTCTGTTCA  
 10 TCTTCTCGTTGGTAGCATTTATGTGAGGCATTGTTAGGTGCATGCACGGTGTAGAATTATTTATCTCTGATGAGTGAA  
 TCTTTGGAGACTTCTATGTCCTAGTAATTCTTTAAATTGCTCTAGTACTGCCACACTGGCTCT  
 TTTGATTAGATTCTCTGCTGTCTGTTCTGCTTAATTATATATATATATATTTTTTTGAGACA  
 GAGTCTTGGCTGTGCCCCAGGGTAGTGCAGTGGTGTATCACAGGTAGTGAACCTTACCTTGGCCTGAGCCG  
 CCTCTCACCTCAGCCTCTGAGTAGCTGAACTGCAGACACGCACCGTACACCTGGCTAATTAAATTCTGGA  
 15 GACAGGGCTTGCTGTGTTGCCAGGCTGCTCAAACCTCTGGACTCAAGGGATCCATCACCTGGCTTCCAAAGTG  
 CTGAATTACAGGCATGAGGCCACCATGTCGGCTAATTCAACACTTTATATTCTTATAGTGTGGTATGCTCTGTTA  
 ACAGCATGTAGGTGAATTCCAATCCAGTCTGACAGTCGGTTAACTGGATAACCTGATTATTTCACTTTGTC  
 ACTAGAGACCCGCTGGCACTGATTCTCACTGGCTGTGATGCTCTGTTCTGACCCCTGTTCTGAC  
 GGTTGCCATGTGCGTTCTGCCAGTGTGTTGATCCTCTCGTTGCCCTGGTCACTGGCATTGCTTATTTCT  
 20 CTTTGCCTAGTGTACCCCTGATCTTTATTGCTGTGTTGCTTTATTGAGACAGTCTACTCTGTCAACCA  
 GGCTGGAGTGAATGGCACAACTCGGCTCACTGCAACCTCTGCTCTGGTCAAGCAGTCTCATTCTCAACCTCA  
 TGAGTAGCTGGATTACAGGCGCCACCAACGCCCTGGCTAATTGAGACAGTCTACCCCTGCTGAGCAATAAGACCC  
 TGCCAGGCTGGCTCAAACCTCTGACCTCAAGTGATCTTAAATGAAGTCTGAAACATTGCTACCCCTGCTGAG  
 25 AGTGTATTTAGCTCTGGCACCCCCCAGCCTGTTGCTGTTCTCTGCTGACTTAGTTCTATCTCAGGCATCTGACA  
 CCCCCACAAGCTAACGATTAAATATTGTTCCGTGAGTGTCTGAGCTTGCCCCGCTGGCTCCACAGTGGATTACAGGTGCAA  
 TTGTTCCCGTCTGCTCTGCTCAGGCCCGCTGGGCTCCATGGCTCTGCTGACTTGCCCCGCTGCTTCTGCTT  
 TTATTGCTGTAACCCAGCTTACCTGCTGGCTCCATGGCTCTGCTGAGCTGGAGCACCACGTGGCCAGCGTTC  
 ACAGATGAAGATGTGGAGACTCACGAGGAGGGCGTCATCTGGCCCTGGTCACTGGAGCACCACGTGGCCAGCGTTC  
 30 CTTAGCCAGTGAGTGACAGCAACGTCGCTGCCCTGGTTGAGCTGGAAACCCAGGCATGCTGGGCTGGTCTG  
 CGCGGTGTCAGTTGAAATCGCGAACCTCGGGTGTGGCGCCAGCTGACGGTGTGCTGGGGAGCTCTGCTTATGTC  
 CTTCCCTCCCTCTGCTGGAAACAGGACAAAGATGAGGCTCGAGGCCGTTGCTGCCAACAGGAGCATGACGTGAGCC  
 ATGTGGATAATTAAATTCTAGGCTGGCGCGTGGCTCACGCCGTAACTCCAGCACTTGGAGGCAAGCGGG  
 TGGATCACAGGTCAGGAGGTCAGGAGGACCATCTGGCAACATGATGAAACCCATCTGACTAAAAACACAAAAATTAGC  
 35 TGGCGTGGTGGCGGGTGCCTGTAATCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTGACCTGGAGTTGGAA  
 GTTGCAGTGAGCCGACATGCACTGCACTCCAGCTGGCAACACAGCGAGACTCTGCTCAAAAAAAAAAAAAAAA  
 AAAAAAAAAAAATTCTAGTAGCCACATTAAAAAGTAAAAAGAAAAGGTGAAATTATGTAATAATTAGATTTACTGAA  
 GCCCAGCATGCCACACCTCATTTAGGGTGTATTGGTGGAGCATCACTCACAGGACATTGACATTGACATTGAGC  
 TTTGCTGCCGATCCCGTGTAGGCTCCGTGCCATCTGCCCTGGACCTGCTGGCTTCCATGCCATGGCT  
 40 GTTGACCGAGATGGTGCAGGTCCGGGATGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGAGTGGCTGGAG  
 TCTGGGATGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGAGTGGCTGGAGGTGAGGTGAGGCTCCAG

GCCCTCGGTGAGCTGGAGGTATGGAGTCGGATGATGCAGGTCCGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGAT  
 TGTGGTGTCTGGATGGTCAGGTCAAGGGTGAGGTCTCCAGGCCCTCGTAAGCTGGAGGTATGGAGTCGGATGATGCA  
 GGTCCGGGTGAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGGTCTGGATGGTCAGGTCTGGAGTGAGGTGCCAG  
 AGGCCCTCGGGTGAAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGGTCTGGATGGTCAGGTCTGGAGTGAGGTGCCAG  
 5 GGTGAGCTGGATATCGGGTCCGGATGGTGCAAGGTCTGGGTGAGGTGAGCTGGATGTGGTCTGGATGGTGAGGTCT  
 GTCCGGATGCTGCAAGTCGGTGTGGTGCGGTCTGGATGGTCAGGTCTGGAGTGAGGTGCCAGACCCCTGCGGTGAGCTGGATG  
 GGGGTGAAGGTGCCAGGCCCTGCTGTGAGCTGGATGTGGTCTGGATGGTCAGGTCTGGAGTGAGGTGCCAG  
 GCCCTCGGTGAGCTGGATGTCAGTGTCAAGATGGTGCAAGTCGGGTGAGGTGCCAGACCCCTGCGGTGAGCTGGATG  
 TGCGGTGTCTGGATGGTGCAAGGTCTGGAGTGAGGTGCCAGGCCCTCGGTGAGCTGGATGTGGAGTCGGATGGTGCC  
 10 GGTCCGGGTGAGGTGCCAGACCCCTGCTGTGAGCTGGATGTGGTCTGGATGGTACAGGTCTGGAGTGAGGTGCC  
 AGACCCCTGCTGTGAGCTGGATATCGGGTCCGGATGGTGCAAGGTCAAGGGTGAGGTCTCCAGGCCCTCGGTGAGCTGG  
 GGTATGGAGTCGGATGATGCAGGTCCGGGTGAGGTGCCAGGCCCTCGTGAGACTGGATGTGCCGGTCTGGATGGT  
 GCAGGTCTGGGTGTGGTCCCAAGGCCCTCGGTGAGCTGGAGGTATGGAGTCGGATGATGCAGGTCCGGGTGAGGT  
 CCAGGCCCTCGGTGAGCTGGATGTGCCGGTCTGGATGGTGCAAGGTCTGGGTGTGGTCCCAAGGCCCTCGGTGAGCTG  
 15 GAGGTATGGAGTCGGATGATGCAGGTCCGGGTGAGGTGCCAGGCCCTCGTGAGCTGGATGTGCTGTATCCGGATG  
 GTGCAGTCGGGTGAGGTGCCAGGCCCTCGTGAGCTGGAGGTATGGAGTCGGATGATGCAGGTCCGGGTGAGGT  
 CACCAAGGCCCTCGGTGAGCTGGATGTGCCGGTCTGGATGGTGCAAGGTCTGGGTGTGGTCCCAAGGCCCTCGGTGAGC  
 TGGATGTGCCGTCTCCCGTCCGGATGGTGCAAGGTCAAGGTGAGGTCTAGGCCCTGGTGGCTGGATGTGCCGT  
 GTCCGGATGGTGCAAGGTCTGGGTGAGGTGCCAGGCCCTGGTGAGCTGGATGTGCCGGTCTGTGATGGTGCAAGGTCTG  
 20 GGGTGAGGTGCCAGGCCCTGGTGAGGTGCCAGGCCCTGGTGAGCTGGATGTGGTGTCCGGATGGTGCAAGGTCCGGT  
 GCTGTGAGCTGGATGTGCCGTCTGGATGGTGCAAGTCGGGTGAGGTAGCCAAGGCCCTCGTGAGCTGGATGTGG  
 GTGTCCGGATGGTGCAAGTCGGGTGAGGTGCCAGGCCCTCGGGTTAGCTGGATGTGCCGGTCTGGATGGTGCAAGGT  
 CCGGGGTGAGGTCAACAGGCCCTCGGGTTAGCTGGATGTGCCGGTCTGGATGGTGCAAGTCGGGTGAGGT  
 CCGGGGTGAGGTCAACAGGCCCTCGGGTTAGCTGGATGTGCCGGTCTGGATGGTGCAAGTCGGGTGAGGT  
 25 CCCGTGTGAGCTGGATGTCTGATCCGGATGGTGCAAGTCGGGTGAGGTGCCAGGCCCTGGTGAGCTGGATGTGG  
 TGCTGTATCCGGATGGTGCAAGGTCTGGCGTGAGGTGCCAGGCCCTGGTGAGCTGGATGTGGATGGTGCAAGTCGG  
 GGTCCGGGTGAGGTCAACAGGCCCTGGTGAGCTGGATGTGGTGCAAGTCGGGTGAGGTGCCAGGCCCTGGTGAGGT  
 AGGCCCTGCTGTGAGCTGGATGTCTGATCCGGATGGTGCAAGTCGGGTGAGGTGCCAGGCCCTGGTGAGCTGG  
 ATGTGCTGTATCCGGATGGTGCAAGTCGGGTGAGGTGCCAGGCCCTGGTGAGCTGGATGTGCAAGTCGG  
 30 TGCAAGTCGGGTGAGGTGCCAGGCCCTGGTGAGGTGCCAGGCCCTGGTGAGCTGGATGTGGTGCTGGATGGTG  
 CGCCAGGCCCTCGGGTGAGGTGCCAGGCCCTGGTGAGGTGCCAGGCCCTGGTGAGGTGCCAGGCCCTGGTGAGGT  
 TGGATATGCCGTGCTCCCGTCCGAATGGTGCAAGGTCCAGGGTGAGGTGCCAGGCCCTGGTGAGGTGCCAGGCC  
 GTCCGGATGGTGCAAGGTCTGGGTGAGGTGCCAGGCCCTGGTGAGCTGGATGTGCCGGTCTGGATGGTGCAAGTC  
 GGGTGAGGTCAACAGGCCCTCGGTGAGCTGGATGTGGCATGTCTGGATGGTCTGGATGGTGCAAGTCGG  
 35 Intron 3 (SEQ ID NO 6)  
 GTACTGTATCCCCACGCCAGGCCCTCTGGTCTCGAAGTCCTGGAACACCAGCCGGCTCAGCATGCCCTGTCTCCACT  
 TGCCTGTGCTTCCCTGGTGTGCAAGTCCTGGCTGGAGGCCAGGGGCCCTGTCACAGGCCCTGGTCAAGTGGATTCTGT  
 CAAGGCTCTGACTGCCTGGAGTCACGTTCTTACTTGAAAATCAGGAGTTGTGCCAAGTGGTCTTAGGGTTGTA  
 AAGCAGAAGGGATTTAAATTAGATGGAAACACTACCAACTAGCCTCCTGCCCTGGATGTGCCGGTCTGGATGGTGCAAGTC  
 40 TCTCTTTTTTTCTTTTGAGATGGAGTCTCACTCTGGTGTGCCAGGTGGAGTGCAGTGGCATAATCTGGCTCACT

GCAACCTCCACCTCCTGGGTTAACCGATTACCAAGCCTCAGCCTCTAACAGCTGGATTACAGGCACCTGCCACAC  
 GCCTGGCTAATTTTGACTTTAGGAGAGACGGGTTTACCATGTTGGCAGGCTGGTCTGAACCTCATGACCTCAGG  
 TGATCCACCCACCTGGCTCCAAAGTGTGGTTACAGGCTAACGCCACCGTGCCAGCCCCGATTCTCTTTAATT  
 CATGCTGTTCTGTATGAATCTCAATCTATTGGATTAGTCAGAGAGGATAAAATCCACCCACTGGGACTCACTG  
 5 CAGGGAGCACCTGTGCAAGGAGCACCTGGGATAGGAGAGTTCACCATGAGCTAACCTCTAGGTGGCTGCATTGAATG  
 GCTGTGAGATTTGCTGCAATGTCGGCTGATGAGAGTGTGAGATTGACAGATTCAAGCTGGATTGATCAGTGAG  
 GGACGGGAGCGCTGGTCTGGGAGATGCCAGCCTGGCTGAGCCCAGGCCATGGTATTAGCTCTCCGTGCCCCGCCAGGC  
 TGACTGTGGAGGGCTTAGTCAGAAGATCAGGGCTCCCCAGCTCCCTGCACACTCGAGTCCCTGGGGGCTTGTGAC  
 ACCCCATGCCCAAATCAGGATGTCAGCAGAGGAGCTGCAGACCTCGTCAGAGGTAACACAGCCTCTGGCTGGG  
 10 GACCCCGACGTGGTCTGGGCCATTCTGCATCTGGGGAGGGTCAGGGCTTCCCTGTGGAAACAAGTTAACAC  
 AATGCACCTTACTTAGACTTTACACGTATTAATGGTGTGCGACCCAACATGGTCATTGACCAGTATTTGGAAAGAAT  
 TTAATTGGGGTGACCGGAAGGAGCAGACAGACAGTGGTGTCCCAGATGCTCTGTCACTACTGGACTGTTGTC  
 CCTGGGGGGCCTTGGAGGCCCTCTCCCTGGACAGGGTACCGTGCTTTCTACTCTGCTGGGCTGCGGCTGGGTC  
 AGGGCACAGCTCGGAGCACCCGGCCCCAGTGTCCACGGAGTGCCAGGCTGTCAGCCACAGATGCCAGGTCCAGG  
 15 GTGGCCGCTCAGCCCCCGTGGCCCATGGTGGTTGGGGAAAAGCCAAGGGCAGAGGTGTCAGGAGACTGGTGGG  
 CTCATGAGAGCTGATTCTGCTCTTGGCTGAGCTGCCCTGAGCAGCCTCTCCGCCCTCTCCATGAAAGGATGTTG  
 CTTCTACCTGGGGTCTGCCTGGCCAGCTGGTGGCTACCCAGTGGCTGTACAGAGGGACAGGCATCTGTG  
 AGGGCATGGGTCACGTGCCAGATCAGCCTGGACCAGGCTCCCTGGTGTGATGGGGACAGTCACCTGGG  
 20 GTTGACCGCCGGACTGGCGTCCCCAGGGTTGACTATAGGACCAAGGTGTCAGGTGCCCCGCAAGTAGAGGGCTTCAG  
 AGGCGTCTGGTGGCATGGTGGACGTGGCCCGGGCATGGCTTCAGCGTGTGCTGCCGTGGTGCCTGAGCCCTC  
 TGAGTCGGTGGGGCTTGTGGCTCCGTGAGCTCCCCCTAGTCCTGTCTGGCTGAGCAAGCCTCTGAGGGCTCT  
 CTATTGCAG

**Intron 4 (SEQ ID NO 7)**

25 GTGGCTGTGCTTGGTTAACCTCTTTAACAGAAGTGCCTTGAGCCCCACATTGGTATCAGCTTAGATGAAGGG  
 CCCGGAGGAGGGGCCACGGGACACAGCCAGGGCATGGCACGGCGCAACCCATTGTGCGCACAGTGAGGTGGCCAGG  
 TGCGGTGCCCTCAGAAAGCAGCGTGGGTGAGGGGAGCTCTGGGGCAGGGACAGGCTCTGAGGACCAAGAAG  
 CAGCCGGGCCAGGGCTGGATGCGCACCGCCAGGACTGCCAGGAGCCCACCGGGCTCTGAGGATCTGGACCTTGGCC  
 30 TCCGTTACGGGGCCGGGACCAGGCCACGACTGCCAGGAGCCCACCGGGCTCTGAGGATCTGGACCTTGGCC  
 CTCTGCACCCACCCCTGTGGCTGGCTGGCGTACCCCGTACATGAGGAGGTGAGGTGGACAGAG  
 GTGTGGCATGAGGATCCCCGTGCAACACACATGCCAGGAACCGGTTCAAACAGGGTCTGAGGAAGCTGGAGGG  
 TTCTAGGTCCGGGCTGGGTGGCTGGGACACTGGGAGGGCTCTCCCTGGTCCCTATGGTGGGTGGC  
 TTGGCCGGATCCACTTCTGACTGTCTCCATGCTGCCCCGCCAG

**Intron 5 (SEQ ID NO 8)**

35 GTGGGTGCCGGGGACCCCGTGAGCAGCCCTGCTGGACCTTGGAGTGGCTGCCCTGATTGGCACCTCATGTTGGGG  
 GAGGTACTCTGGTGGGCCAGGGAGTGCAGGTGACCCCTGTCAGTGTGAGGACACACCTGGCACCTAGGGTGAGGC  
 CTTCAGCCTTCTGCAGCACATGGGCCAGTGTGCAACCTGACTGCCGGCTCTATTCCAAGGAGGGTCCACTG  
 GATTCCAGTTCCGTAGAGAAGGAACCGCAACGGCTAGCCACCAAGGCCCGGTGCCCTGCACCCAGTCTGAGCCAG  
 40 GGGTCTCCTGTCTGAGGCTCAGAGAGGGACACAGCCCGCCCTGCCCTGGGTCTGGAGTGGTGGGGTCAGAGAGAG

AGTGGGGGACACCGCCAGGCCAGGCCCTGAGGGCAGAGGTGATGTCAGTTCTGGCTGGCCACTGTCAGTCTCCCTGC  
CTCCCACTCACACAG

### 5'-Bereich Intron 6 (SEQ ID NO 9)

5 GTAAAGGTTCACCGTGTATAGTCGTGTCAGGATGTGTCTCTGGGATATGAATGTCTAGAATGCACTCGTGTCTGTG  
ATGCGTTCTGTGGGGAGGTACTTCCATGATTACACATCTGTATAGCGTGTGCGCACGTGTGTGCGCTGGTGCAT  
GTATCTGTGGCGTGCATATTGTGGTGTGTGTGTGGCACGTGTGTCCATGGTGTGTGCGCTGTGGTGCATG  
TGTGTGTGTCTGTACACGTGCATGTTCATGCTGTGTGCTCATGTCATGTGTATGTGCTATTGTGGTGTGTGTGCAT  
GTGTCGTGACATATCGTGTCTATGGCATGGGTGTGTGCGCCCTGGCCTACTCCCTCCTCCAGGCATGGTCC  
10 GCACCAATTGTCCTCACGCTCTGGGTGCTGGTTGGGAGCTCACATTAGGGTCCTCACCTAGCATGGTGCCCC  
GTCTGTACAGGGCTGGCTTGAGACTGTAAGCCAGGTTGAGAGGAGTAGGGATGCTGGTGTACCTTCCTGG  
CCCCCTGGCACCCCCCAGGACCCCACTGTCCTATGCCCTGGCCTGGCCTGGCAGGGTGAAAGGGCCCTGGCTGGTCCACCC  
GGGACACACTCCTCCCAGAGCGGCCGGGGCTTGGGCTCGCAGGGTGAAAGGGCCCTGGCTGGTCCACCC  
15 AGTGGTCATGAGCACGCTGGAGGGTAAGCCCTAAAGTCGTGCCAGGCCGGGTGCAAGGGTAAGAAGTATCCCTGG  
GCTTCGGTCTGGGGAGAGGCACATGTGGAAACCCACAAGGACCTTTCTCTGAATTCTGGACT

### 3'-Bereich Intron 6 (SEQ\_ID NO 10)

20 TGTGGGATTGGTTTCATGTGGGATAGGTGGGATCTGTGGGATTGGTTTATGAGTGGGAAACACAGAGTTCAG  
GCGAGCTTCTCCTGTAGTGGGCTGCAGGTGCTCAAACAGCTTATTGAGGAGACCATATCTCCCTTGAACTATGGT  
CGGGTTATAGTAAGTCAGGGTGTGGAGGCCTCCCTGGCTCCCTGTGTTCTGTTCTCCACTCTGGGTCGTGGT  
CTCTGTGGTGTGGCCGGTGGCAGGGCTCCAGGCTCCTGGTGTGTCATTGGCTGGATGTGGCCCTGGCTACCGT  
CCGTCTTGGAAATCCCCCTCGAGTTGGAGGTTCTTCTTCTTTCTTCTTTCTTTCTTTCTTTGATAACAGA  
GTCTCGCTCTTTTGCCCAGGCTGGAGTGGTTGGCGTACTGCAACCTGTGCTTCTGAGTTCAAGCA  
ATTCTCTGCTCAGCCTCCAAGTAGCTGAATTATAGGCCACACCATGCTGACTAATTGGTAAATTAGTAG  
25 AGACGAGGTTCTCCATGTGGCCAGGCTGGTCTGAACCTGACCTCAGGTGATCTCCCACCTGGCCTCCAAAGT  
GCTGGGATGACAGGTGTGAAACGCCCGCGCCGGAGACTCGCTTCTGCACTGGTGGAGATCTGAGCTAGCTG  
CTCGACCCCTGGTGTGACAACCTCCCTTCTCCAGGTCTCGTAGGGTCTTCCATTCTGACTCTCTCA  
CAGAAGAGTTTCACGTGTGTGATTTCGGCTGTTCTCGTAATTGGTGTCTGCTGTTATGATGGCCCTCTCCA  
TTTCCTTCTAGGCTTGTATTGGTTTCTGGCTCTGAAGGAAAGTTGATATTGGATGTTGAACCTTCTT  
30 TCTAAACAAGCATCTGAAGTGGCTTCTCTAAAGCAGGGATCCGAGGCCCTGGCTGGAGTGGCACGGTCT  
GGGGCCTGTTAGGAACCCGGCGCACAGCGGGAGGCTAGGTGGGTGTGGGAGCCAGCTTCCGCTGAGCCCCGGCC  
TCTCAGATCAGCAGTGGCATGGGTGCTCAGGGCGCACACACCCACTGAGAACTGTGCGTGAAGGGGTCTAGATTCT  
GTGCTCTTATGGGAATCTAACTGCTGATGATCTGAGGTGGAACCGTTGCTCCAAAACCATCCCCCTCCCCACTGCTG  
TCTGTGAAAAATCGTCTCACGAAACCACTCCCTGGTACCAATGGTTGGGACCCCTGTCTAAAGACCTGCTCA  
35 GCAGCCTCTGTCAGTGTGATATATTGGCTTCTGTGTTGAGTCCAGAATAATTACGGATTCTGTGATGCTTCCG  
CGACCTCAGACCCATGGGCTATTGTGGCGTGTGCTCTGGTTGGGAAGGGTGCAGGCCCATGTAACCTTCT  
GTTACTGCTTCCAGGTTGGTCTCAGGGTTGAATCGTACTCGATGGTTTAGCCACGGCCCTGCCAGCTCTG  
GGGGCTGGGAACATGCTGAAGCACAGACTACCCTGGCGTCTTGTGCTCACAGCTGAGGCCCTGTGCTCG  
TGTGAGTGTGTCACGTGCTCACACCTGCTTGGGACGCAAGGGCTTAGCAGGTCCGTAGTAAATGACAAGC  
40 GTCTGGGGAGTCTGCAGAAATAGGAGGTGGGGTGGGGTCTCTCTCCCGCTCTGAGACTCTTCTGCTGTGCT

GTGGCTGCACCTGCATCCCTGCAATCCCTCCAGCACTGGGCTGGAGAGGGCCGGAGCTCGAGTCGCACATTGTCACCGT  
GACTGTGGATGGCAGTCGGTCACGGGGTCTGATGTGGTACTGTGGATGGCGGTCTGATGGGCTGAGTCGGTCACTGGGCTGATGTGTG  
GTGACTGTGGATGGCGGTCTGGGGTCTGATGTGGTACTGTGGATGGCGGTCTGGGCTGATGGGCTGAGTCGGTCACTGGGCTGATGTGTG  
ATGGCGGTCTGGGGTCTGATGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGGTACTGTGGATGGCGGTCTGGGCTGAGTCGGTCACTGG  
5 GGGTCTGATGTGGTACTGTGGATGGCAGTCGTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGGTACTGTGGATGGCGGTCTGATG  
TGGTGA  
GTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGGTACTGTGGATGG  
CGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGTCA  
CAGGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGTACGGTCA  
10 GGGTCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTG  
TCTGATGTGTGGTACTGTGGATGGCGATCGGTACAGGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCT  
GATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGT  
GACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGG  
GGCGGTTGGTCCCGGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCAG  
15 TCGTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCT  
TCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGT  
GGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCT  
GACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCT  
GGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCT  
20 GTCACAGGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCT  
GGTCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCT  
GTGGTACTGTGGATGGTACCGTCACAGGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCT  
ACTTTGCGTCTCGGCCCCCGGCCCGTTCACAGGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCTC  
GGCTTGGCCGAGGTCCACACGTCTGATCGGAAGAACAGAAGTCCCAGGGCTGAGTCGGCAGGAGTCCCAAGGTCTGATGTGTGGTACTGTGGATGGCGGTCTGGGCT  
25 ATGCCCTCTCTGCGCGAG

**Intron 7 (SEQ ID NO 11)**

GTCTGGGCACTGCCCTGCGGGGTTGGGCAACGGACTCCAGCAGTGGGCTCCCTGGCAATCACTGGGCTCATGACCC  
 GACAGACTGTTGGGCTGGGGGCACTGGGGGAATGAGCTGTGATGGGGCATGATGAGCTGTGCTGGCGAAATC  
 30 TGAGCTGGGCCATGCCAGGCTGCGACAGCTGCTGCATTCAAGGCACCTGCTCACGTTGACTGCGCGGCCCTCTCCAGTT  
 CCGCAGTGCCTTGTTCATGATTTGCTAAATGCTTCTGCCCAGTTGATCTGAGGCCAAAGGAAGGTGTC  
 CCCCCCT  
 CCTTTAGGAGGGCAGGCCATGTTGAGCCGTGCTGCCAGCTGGCCCTCAGTGTGGGCTGAGGCCAAAGGAACG  
 TGTC  
 35 CCCCTCTCTTGTAGGAGGACGGGCCGTGTTGAGCCACGCCCGCTGAGCGGGCTCTCAGTGTGGGCTGTCA  
 CGT  
 GGCCCTGTGCCCTTGCAAGATGTGGTCTGTCACGTGGCCCTGTGGCTTTGCAAGTCCTGAGCACTGGCT  
 CGGAG  
 TCTAGGGACAGTCGTCTCCACCGCATGAGGCTCAGAGACCTCTGGCGAATTCTGGCTCCAGGGTGGG  
 GTGGCCTGGGCTGCTGGGACCCAGACCCCTGTGCCCGCAGCTGGGAGCAACTCTGGATCACATATGCC  
 40 CGGTGGGCTGTGTTGGGTGAGCCCAGCTGGACCCACAGGTGGCCAGAGGAGACGTTCTGTGTC  
 ACACACTCTGCC  
 GCCCATGTGTCTGCAGAGACTCGGCCGGCAGCCACGATGGCCCTGCATTCAAGCCAGCCCGCA  
 AACACTGACCCCCAAAGGGACGGAGGCTTGGCCACGTGGCTCTGCTCAGCACCCACCG  
 AACACTCCATGTC  
 TCTCCCGTGTGCTTCGCAAG

## Intron 8 (SEQ ID NO 12)

5 GTGAGTCAGGTGGCCAGGTGCCATTGCCCTGGGGTGGCTGGCGGGCTGGCAGGGCTCTGCTCACCTCTCTCCGCC  
 CTTCCCCACTGNCCCTCTGCCCGGGGCCACAGAGTCCTCTTCTGGGCCCCGCCCCCTCCGGCTCTGGCTGCAGGC  
 TCCCGAGGCCCCGGAAACATGGCTCGGCTTGGCAGCCGGAGGGAGCAGGTGCCACAGGAGCTGGAAATGGCAAGC  
 GGGGTGAGGTTGCTCTCGCTGGAGGACAGGGGGGGGGGTGTCTGGGTCAAGGTGCGCCGAGCGTTGAGCCT  
 GCAGCTGTCAAGCTCCAAGTTACTACTGACGCTGGACACCCGGCTCTCACACGCTTGTATCTCTCTCCGATA  
 GAAAAA  
 GGATTTATCCGATTCTCATCCCTGCTGTGACCCCCCGAGGGCGGGCTCTCTCTGTGACTAGATT  
 CCCATGGAAAAGTGCAGGGTTGACCGTGTAGTTGCTCCTCTGGGGGCTGTGGTGGCCATGGGCAGGCGGCC  
 10 GAGAGTCGCCACACAGCCACTGGGTGAGGCCACTCACGGTGTAGAGCCACAGTGCCTGGTGCACATCACGCT  
 CTGGATTTAAGTAAAACACACACCTCCCGCAGGCATCTGCCTGCACCCCTGTGTGCTGGCTGGGAGAGTGGTAC  
 GCAGGAAATTGTCGACACTCAAGGTCAAGGTGACAGCTGGAGCTTTATTTAAAAATAACTATTAAATTG  
 CATTATAAGT  
 15 AATCACTAATGGTATCAGCAATTATAATATTAAACTATAATTAGAAAATTAAAGTAGTACACACGTTCTGGAAAA  
 CACAAATTGACATGGCAGCAGAGTGAATTGGCCAGGGACAGTGTGCACATGTGTGAAGCGGCCAGGCCAC  
 AGAATTGCTGACAAAGTCACCTCCCCAGAGAAGCCACACGGGCTCTCGTGGTGTGAATTATAAGATGGATC  
 AAGTCACGTACCGTCCACGTGTCGGCAGGGCTTGGGAATGTGAGGTGATGACTGCGCTCTCATGCC  
 GACAGACAGGA  
 GGTGACTGTGCTGCTGCTGCTAGGACACGGCAGGGCGAAGCTCTAGTCCCCATGTTGGCCAGTTGGC  
 20 ATAAAAACGTCTCAAAACCTGTTGCCCAAAACTAAGAACAGAGAGAGTTCCCATCCC  
 ATGTCACAGGGCGTA  
 TCTGCTTGCCTGACTCGCTGGCTGGCGACTCTAGAGTTGGCGGTGCTCTGTGCAAAAGTGCAGTC  
 GCCCATCACTGTGATATCTGACCGCAAGGAAGGCCCTTTCTTTCTTTCTTTGAGACGGAACTGCA  
 CTGTTGCTGCTGGCTTGGCTTGAGTGCAGTGGCGCATCTCAACTCACTGCAACCTCCGCT  
 25 CTGCTCAGCCCTCCGAGCAGCTGAGATTACAGGCACCCACCCCTGCGCTGGCTAATTTGTATTTAGAGAG  
 GGGTTTTGCACTGTTGGCAGGCTGGCTCGAACCTCTGACCTCAGGTGATCCACCCACCTCGGCT  
 CCAAAGTGTG  
 GGATTACAGGTGTGAGCCATCACGCCAGGGAAAGCCTCTTTAAAGGTGACCACCTATAGCG  
 30 CTCCCGAAAGCTCTTCTGAGCTTGGCTCTGAGCTTGGCTCTGTGGCTCTGGGATGGCTGAGGG  
 TCGCGTGGCAGCCATGCCCTGTTGGAGAGTTCTGCTCTCGTGGCTAGCTGAAACTAGGG  
 CCCTTGGCATCTTGTGAGAGTTCTGCTCTCGTGGCTAGCTGAAACTAGGG  
 GCAGCGGCTACATGTAGGGTCATGAGTCTTCAACGGTGGACAAATTCT  
 35 30 TAATTCCGGTCAAGTGTGTTCTGTGAAATAACTCTAAGATTAAAGAAACCTTA  
 AGAACCTTGATGATTC  
 AGAGCAAGGATGTGGTCACACCTGTGGCTGGATCTGTTCTGACCC  
 CAGTCAGTGGAGAGTGGCTGAGGTGTTGGCTGGCTGAATGGTAGAC  
 GTGCTGTTGTGAGTGTACG  
 TCTGTGCTGTTGGCTGGCTGGTTGAGTGTACG  
 CATGTCAGCACATGCC  
 CCCAG

## Intron 9 (SEQ ID NO 13)

GTGAGGCCTCTCTTCCCCAGGGGGCTGGGTGGGGTTGATTGCTTGT  
 40 GATGCAATTGAGCTGAGGACAGACAAGGGTGCAGCCCTCTGGTATGAAGCC  
 GCACGGAGGG  
 TTGCAAGCCTGAGGACTGCGGGCTCCACCGCAGGCTCTGTCAGCG  
 GGGCATGTCAGAGGCTCAGGGCTCAGGGCTCAGGGCTCAGCAGGCC  
 GAGGGCCGCTGCC  
 CTGCA  
 GATGAGCATGTGAATTCAACACCGAGGAAGCACACCAGCT  
 GTCACGT  
 CACCCAGGTT

CGTTAGGGCTTGGGGAGATGGGGCTGGTGAGCCTGAGGCCACATCTCCAGCAGGCCCTGACAGGTGGCTGGA  
 CTGGGCCCTTCAGCCCATTGCCACTTGATGGGTCTACACCCAAAGGACGCACACACCTAAATATCGGCC  
 AACCTAATGTGGTCAACTCAGCTGGCTTTATTGACACGAGTTACTTTTTTTAATACTTAAGTTCTAGGGTAC  
 ATGTGCACGACGTGCAGGTTAGTTACATATGTATACTATGATGCACTGTTGGTGTGCTGCACCCATTAACTCATCATTACA  
 5 TTAGGTATACTCCTAAATGCTATCCCTCCCACCTCCCCCATCCATGACAGGCCCTGGTGTGATGTTCCCCACCCG  
 TGTCAGGAGTGTGATGGTTCCAGCTTCGTCCATGTCCTACAAAGGACATGAACCTCATCTTTTTATGACTGCATAGTATT  
 CGCTGGTGTATATGTGCCACATTTCTTAATCCGTCTATCATGATGGACATTGGGTTGGTGCAGTCTTGCTACT  
 GTGAATAGTGCAGCAATAAACATACGTGTCATGTCATTTATAGCAGCATGATTATACTCTGGTATATACCCG  
 10 TAATGGGATGGCTGGGTCAAATGGTATTCTAGTCTAGATCCTGAGGAATCACCACACTGTCTCCACAATGGTGA  
 CTAGTTACCTCCACCAACAGTGTAAAGTGTCTGGTGTGGAGGGATGTGGACAGCAGTTATTTTTATGAAAA  
 TAGTATCACTGAACAAGCAGACAGTTAGTGAAGGATGCGTCAGGAAGCCTGCAGGCCACAGCCATTCTCGAAGAC  
 TCCGGGTTTCTGTGCATCTTGAAACTCTAGCTCAATTATAGCATGTACAGTGGATCAAGGTTCTCTCATTAA  
 GGTTCAAGTTCTAGATTGAAATAAGTTATGTAACAGAAAACAAAATTCTGTACACACAACCTGCTCTGGGATTGGA  
 15 GGAAAGTGTCTCGAGCTGGCGCACACTGGTCAGGCCCTGGACAGGATACCTCTGGCCATGGTCATGGGCGCTGG  
 GCTTGGGCTGAGGGTCACACAGTCACATGCCAGCTCTGTGGATAGGATCTGGGTCTGGATCATGCTGAGGACC  
 ACAGCTGCCATGCTGGTAAAGGGCACACGTGGTCAGAGGGGCGAGGTTCCAGCCCCAGCTTCTTACCGTCTCG  
 TTATTTCTCTAAGAGTGTGAGAAGTGGGGCGCCGTGATGGCCTCGTCTCAGCTGGCACAGAATTGACAA  
 GCTGATGGTAAACACTGAGTACTTATAATGAATGAGGAATTGCTGTAGCAGTTAACGTAGAGAGCTGTCTGGAAA  
 20 GAAATTAAAGTTTCTATTAAACCGCTTGGAGAATGTTACTTTATTGCTGTGAAATTGTTGACATTCACTGCCC  
 TCGTAGACAGATACTACGTAAAGTTAACCTGCTGTATTTCCCTTATTTAG

**Intron 10 (SEQ ID NO 14)**

GTGAGGCCGTGCCGTGTCTGTGGGACCTCCACAGCCTGTGGCTTGAGTTGAGCCCCCGTGTCTGCCCTGG  
 25 CACCGCAGCCTGTCCTGCCAAGTCCTCTCTGCCGGTCTGGATCCGCAAGGCAGAGGCCCTGGCGTGCACC  
 CAGGCCCTGGGGCGCACGGGACCTTGGGAGGGAGTGGGTACCGTGCAGGCCCTGGCTCGAGACGCCACCGAGT  
 ACACACGTGGTGAAGTCAGGCCGTGACCTGGCTCTGTCTTTGGAAAGTCAGAGTGGCGCTCCTGGGCCCCAG  
 TGAGACCCCCAGGGACTGTGACAGGCCCTGCAGGGCCGAGGCCAGCCTCCCTCCAGGGTGCACCTGAGCCTGCCA  
 GAGCAGGAGCTGCTGAGTGAAGCTGGCCACAGCCTGGCTGCCGTACGTTCTGGCTGGGGTTTTGGATGGGG  
 30 AGAATTGGATTGCTGAGTGTCTGTCTTGAAACCACGGAGATGGCTAGGAGTGGTTCTAGAGTTGATTTGTGAAT  
 CAAACTAAATCAGGCACAGGGACCTGGCTCAGCACAGGGATTGTCATGTGGCTCCCTCAAGGGCCCCACAG  
 AGCCGGTGGCTTAAAGTGCATTGACAGGGAGCAGAGAACCTGAAAGCTGAAAGGGACCCCTCAGAAAATG  
 TGGCCGCCAGGGTGGTTCAAGGTGCTTGTGGCTGTGTTGTGAAAACCCATTGGACCCCTCCAAGTCACCC  
 TCCAGGTCCACCCCTCCAGGGCCCTGGCTGGGTATGCCCTGGCTTGTGCCAGGCCAGCACAGCAGGC  
 35 TGTGCACATTAAATCCACTAAGATTCACTCGGGGAGGCCAGGTCCAAGCAACTGAGGGCTCAGGAGTCTGAGGCT  
 GCTGAGGGAGAGCAGACGGGAACGCTGCTCTGTGAGTGGCAAGTCTGAGGGTGTGGCCAGGGAGGTGGCTCAGA  
 GTGTATGTTGGGTCCACCGGGGGCAGAACACTCTGTCTGTGATGAGTGGCAGCCATGTAACAGGAAGGGTGGCAG  
 GGAGCTGGGAATGCACCAAGGGAGCTGGCAGCTGGCCAGGTTCCAGGGCAGGCCACAGGAAGGGCAGGGGGACGCC  
 40 GGGGCCACAGCAGAGGCCAGGAAGGGATGCCAGGCCAGAGCAGAGCTACGGGCACAGGGGGCTCCCTG  
 AGCTGGGTGAGCGAGGCTCATGACTCGGCCAGGGAACCTCCTGACGTGAAGCTGACGACTGGTGTGCCCCAGCTCACAG

CCCAGGCCAGGTCCCGCGCTGAGCAGGAACCTCAGAACCCCTCCCTTGTCTAAAGCACAGCAGATGCCCTCAGGGCATCTAGGAGAAAACAGGCAAAGTCGTTGAGAAACGTCTAAAGAAGGTGGGATGGTGGCAATTCTTGTCCAGATTTCAGTTAGCTGCCCGGACACAGATGAGTCTATAACGGGATTGTTGTCATGGGACACATGAGATGGACCATCACAGAGGCCACTGGGGCTGCACCTCCCATCTGAGTCCTGGCTGTCCCGGGTCAGGCCAGGTTCTGCTACCTACCTGTCTCTGCCGGGAGACAGGGAAAGCACCCCGAAGTCTGGAGCAGGGCTGGTCCAGGCTCCATCAGAGTCCTGCCAGGCCAGCACCCCTGCTCAAATCACCACCTCTGGGTTTCCAAAGCATTTAACAGGGTGTCAAGGTTACCTCCTGGGTGACGGCCCCGATCTCTGGGGCTGACATTGCCCTCTGCCCTAG

Intron 11 (SEQ ID NO 15)

10	GTGAGGCCACCTGGCCGGAAAGTGGAGCCTGTGCCGGCTGGGGCAGGTGCTGCTGCAGGGCCGTGCGTCCACCTCTGT TCCGTGTGGGCAGGCAGTCCAATCCAAAGGGTCAGAGGCACAGGGTCCCCCTCGTCCCATCTGGGCTGAGCAGA AATGCACTTTCTGTGGGAGTGAGGTGCTCACAAACGGGAGCAGTTCTGTCTATTGGTAAAGGAAATGGTGAC CAGACCTGGGTGCACTGAGGTGCTTCAGAAAGCAGTCTGGATCCGAACCCAAAGACGCCGGGCCCTGCTGGGCTGAGT CTCTCAAAACCGAACACAGGGCCCTGCTGGCATGAGTCCCTCTGAACCCGAGACCTGGGCCCTGCTGGGCTGAGT 15 CTCTCCGAACCCAGAGACTTCAGGGCCCTTGGCGTGAGTCTCTCCGCTGTGAGCCCCACACTCCAAGGCTCATCCAC AGTCTACAGGTGCCATGAGTTCATGATCACGTGACCCATCAGGGACAGGCCATGGTGTGGGGGGCTCTACAA AATTCTGGGTCTTGTCCCCAGAGGCCAGAGCTCAAGGCCGCTCAGGCTCAGACACAAATGAATTGAAGATGGA CACAGATGCAGAAATCTGTGCTTTCTTATGAATAAAAGTATCAACATTCCAGGCAGGGCAAGGTGGCTCACACCT ATAATCCAGCACTTGGGAGGCCGAGGTGGTGGATCACTTGAGGCCAGGAGTTGAGGCCACCTAACCAACATAGT 20 AAATTCCATTCTACTTAAAAAATACAAAAATTAGCCTGGCTGGCACACGCCCTGTAGTCCCCCTATGCCGGAGGC TGAGGCAGGAGAACATTTGAACCCAGGAGGCAGAGGTGAGTGCAGGCCAGGATCACACCAACTGCACTCCAGCCTGGCA ACAGAGTGAAGACTTCATCTTAAAAAAGTATCAGCATTCCAAAACCATAGTGGACAGGTGTTTTTATTG TGTCCCTCGATAATTTACTGGTGTGCTAGAGGCCGAACTGGGGTGCCTCTGAAAGGCACACCTTCTGG GAAGAGAAATAAGTGGTGAATGGTTAAACAGGGTTAACTGGGTCTGCTGAGTTAACAGTCCAGATC 25 TGGACTTTGCCCTTCCAGAAATGCTCCCTGGGTTGCTCATGGGGAGCAGCAGGTGGACACCCCTCGTATGGGG GAGCAGCAGGTGCAGACGCCCTCATGATGGGGAGTGGCAGGTGCAGACACCCCTGTGCATGGGCCAGCATGCCCTG TTGCAGCTCCCTCCCCACAAGGATGCCGGTCTCCGTGCTCCACAGTCCCTGCTTCCCTCAACGCCCTACCTGGT CTGGCCTCCACTGGCTTGTCTGCATGATTCCACATTCTGGCTCCAGCACCTTCTGCTCTCCAGGCACCT GCAGTGTGCCATACCAGTCAGTGTGAACTGTCCACTGTTATTGCTCCCCATGAAATGTATTAGGACAGGG 30 ACCCCTGGTCCAGCCTCTGGCACAGCATCAGTGAATGTTATTGAAGGACAAAGGACAGACAAACAAATCAGGAAATGG GTTCTCTAAACACATTGCAAAGCCACAGGGCTAGTGCAGGATGGTGGCATCAGGTCTCAGATGTGGTCCAATG CCAGAAATATTCTGTGCTCCAAAGGCCACTGGTCAGAGTGTGCTGAGGGTGGCTCTAAAGCTCAGCAGTGGG GCAGTGGTTCGCCATACTCAGGGTGAACTCACATCCTGTGCTGAGTATACAGCAGAGGCTGAGGGCATCTGGG GAAGAAAACAGGCAAATGATTAGAAAAGTGAAGGAAAGTGGTAAGATGGAAATTCTTGTCCAGATTAGTC 35 TCCCCAACACAGCTCAGATGGTGAATGTGTCAGAACAGTGTGGACAGAACAAATAGAACAAACGAAGCCCTATCT CAGAAACAGTGTGTTAATGTGGTATGTGGCACAGCTGATGGAAAAGAGAGTGTGTTGTAATTCTGAGAAAAC GACTGGAAAGCAAATAAGTTGTCTTACAGCATATACAGAGCAGATTCTAGGTAGAAGAGGGACACATGCAAACAC ACCAGCAACAGAAATAACAAACAAAGCTCAAAGGAAAGGGAGGTGAACGTTCCCTGTTGGTGTGAGGGAGGACAC AGGGAGGCGGATGAAACCGTAGTGGCAACGGGATTGCTTCACTGCAGAGAAAACCTAGCTTGCCTGAGCCACAGTGA 40 ATGGCCATTCCCTGGAGCGCTTGTGCACTGTGATTATTAGGCGCCCTGTGAGGTTCTGCACTTCACTCCCTCACTT
----	---

5            GTTCTCCTAACCAACCTGAGAGGTAGAGGAGGAAGGCTCCAGGGAGCAGCCGCCCTGGTCACCCAGCTGGCAAAGGGC  
 ATGCATGATTGCAAGCCTGGCCTCTGCTCCGGGCCCTGCTCTGCCAGGACCCACACAAGTCAGACCCATAGGCTC  
 AGGGTGAAGCCGGAGCCAAAGGTCGTGGGGATGGCTGTGAAGAAGAAATGGACGTCTGATGCACACTTGGAAAGGTC  
 CTACCAGCAGCGTCAAAGAAATGCATGTGAAACTGACAGCGAGACCCATCCCTCAAAGAAACGCACGTGAAACTGATGGC  
 GAGACCTGCCCCATCCCTCATGCTGGCTCTTCTGGCTTGCAGGAGGCCAGCATCAGGTTGAGGCAAGCTGGAAAG  
 ACTTTCTGAAAGCAGCTGTTGCAATGGAAGTCCTACAATGTCCTGTCTTCCAGTAATTCACTTCTGAAGTGA  
 CCAGACATTATCACGGGTCTTATTACCATTCAGTGTCCAGGCAGGGGACTTGCACAGCAAGTCACGAACCTGCC  
 CAAATACAGGGCTAAGGAGATATTATGCATCACAAACTGCTCTGCATTAAACATTITCAAAGAATTITGAAGAAT  
 GTTTAATGGCACAAACGTTATTCAATGTAGCAGTGTCAAAGCTGGATGTAAAGAACACACCCAGGAGCCTGCCG  
 10            TGAATGTCATGTGTGTTCATCTTGACATGGACATACATGGCAGTGTGAGTGGTGTGAGGGCTGGAGGACATCGGTGG  
 GATGCCCTCATCCTGCCCTCTGGAGACACCATGTGTGCCACGTGCACTCACTGGAGCCCTGTTAGCTGGTGCACCTG  
 GCTCTCCATCCCTGAGATTCAAACACAGTGAGATTCACAGGCCAACTCAGTGTCTCCACAAAAACCTGAGTCAC  
 ACCTGTGTTCACTCGAGGGACGCCGGAGGGCTCCACAGTTATTATGTGTTTGGCTGAGTTATGTGAGAC  
 TCATCAGGGCAGATGATGAGTCACAAACACGGCGTGGCAGGTTGGATAACTCAACATCACTAGCCAGGTCTGGT  
 15            GAGTTGGTCATGCAGACTGGATGGCATGTAGCAATTGGAGTCCATGGAGTGACCAACCCAGGCCCTGGCTGCAGC  
 GCATGCCCAAGGAGAACAGGAAGCGGGAGGAAGGAGGAGGCTTTGGAGCAAGCTTGCAGGAGGGGCTGGTGT  
 GGGCAGGCACCTGTGTCGACATTCCCCCTGTGTCAG

**Intron 12 (SEQ ID NO 16)**

20            GTGAGCAGGCTGATGGTCAGCACAGAGTCAGAGTCAGGAGGTGTGCGCAAGTATGTTGTTGTTGCGCCGCT  
 GCCTGCAAGGCTGATGGTACTGGCTGCACTGAAGAGTCACATGTACGCATATAACAGTGAGCACATACATGTCAT  
 GTGTTGACATGAAGGCATGGCAGTGTGTCACAGGTGTCAAGGGCACAAGTGTGTCACATGCGAATGCACACCTGACA  
 TGCACTGTGTTGTCGACAGTCGTGTCAGGTGCACTGGCTGAGGTGCACTGGCTGTGGGTGTGCACTGTGAGTAGCATGT  
 GCACATAACATGATTGAGGGTCTCGTGTTCACCCCTGAGGTCTCAGCACCAGTGCCACTCCTAACGGATGAGAC  
 25            GGGGTCCCAGGCCCTGGGGCTGAGGCTCTGAAGCTGCAAGCCCTGAGGGCATTGTCATCTGGGATCCGCGTCCACT  
 CCTCTCCTGTTGGCTCTGTGTCACCTCCCTCTCTGGGCTTACATCCACTCCACTCCCTCTCTCTGTGGC  
 ATCCCGTCACTCCCCCTCTGTGGGATCTGCGTCCACCTCCCTCTGTGGGATCTTGGCTCACTCCCTCTGTGGC  
 GGTTCCTCTGTCTGGGAGGCCCTGGGGCAGGCAAGTGAACAGAGTCCTGACTGCCAGGGTGGTGTGCACTG  
 CCGGGTGAAGGGCCAGGGGATTCACTGGGAAGAGGGATAGTTCTGTCAAATGTTCTCTTGTGTCATCTGA  
 30            ATGGATGATAAAGCAAAAGTAAAGCTAAATCCAGAGGGTTCTACCGTTCACTCTTCTGGGACTCTAG

**Intron 13 (SEQ ID NO 17)**

35            GTGAGCCGCCACCAAGGGGTGCAAGGCCAGGCCCTCCAGGGACCCCTCCGCGCTCTGTCACCTCTGACCCGGGGCTTACCT  
 TGGAACTCTGGTTTAAGGGCAAGGAATGTCTTACGTTTCACTGGTAGGGTAGGAGGCCGTGGCCCCAGGTGCCCCACTGTGCTGT  
 GCTCTGTGCAAAGCACCTGTTCTCCATCTCTGGTAGGGTAGGAGGCCGTGGCCCCAGGTGCCCCACTGTGCTGT  
 GCACTGGCCGTGGGACGTCACTGGAGGCCATCCCAAGGGCAGCAGGGGATGGGTAAGAGATGTTATGGGGAGTCTTAG  
 CAGAGGAGGCTGGGAAGGTGTGAACAGTAGATGGGAGATCAGATGCCGGAGGAATTGGGGTCTCAGCAAAGAGGGC  
 GAGGTGGGTGCAAGGTGAGGGTCGCTGGCCACCCCCGGGAAGGTGCAAGCAGAGCTGTTGCTCCACACAGCCGGCA  
 GCACCTGTGCTCTGGCATGGCTGTGTCCTGGAACGTTCCCTGTCTGGCTGGTCAAGGGGTGCCCTGCAAGAATCG  
 40            ACAACTTATCACAGAGGGAAAGGCCAATCTGTGGAGGCCACAGGGCAGCTCTGCTGGAGTCAGGGCAGGTGGC

ACAAGCCTGGGGCTGTACCAAAGGGCAGTCGGGACCCACAGGCCGGGCTCCACCTCACAGGCCCTCCGAGCCACTG  
 GGAGCTGAATGCCAGGAGGCCAAGCCCTGCCCATGAGGGCTGAGAAGGAGTGTGAGCATTGTGTTACCCAGGGCG  
 AGGCTGCGCAATTACCGTGACACTTGATGTGAAATGAGGTGCGTCATCGTGGAAACCCAGCAAGGGCTCACGGGA  
 GAGTTTCCATTACAAGGTGTAACCATGAAATGGTTTAACCCAGTGTGCGCCTCATGCTCTGGCAGGGAGGGC  
 5 AGAGCCACAGTGCATGTTACCGCTTGACCAAGCTCCAGGGCTTGGGACCAAGGCTGTCTCAGTTCCAGGGTGCGTCC  
 GGCTCAGACGCCCTCCTCTGCTTCTCTGCTCAAATCTCCCTGTTGACATCTCCCTGACCGTGCTGG  
 CCCTCGTGCAAGCTGCTTGACTCTTCCGAAACCCCTGGGTTGCTGGATACAGGTGCGACTGAGGACTGGAGGTGT  
 CTGACACTGTGTTGACCCCAGGGCCAGCTGGCGTGTGGGCTCCGGGATGAGGTGAGGGAGTTC  
 CAGGTGAAAACCTGGAAACTCCAGGGCCATGTGACCTGCCACCTGCTCTCCCATATTCA  
 10 CAGTCAGTCTCTGGGAACTCCAGGGCCATGTGACCTGCCACCTGCTCTCCCATATTCA  
 CCATGTGGGACCCCTGGTAGCTGCTGTTGAGGGTAGCCCTGAGGAGGGCAGATGCGATGGGCCACGGGCC  
 AACACAGAGTCAGGCACGTGGAAGGCCAGGAATCCCCTCCCTGAGGAGGGCTGGGAGAAGCTGGGAG  
 ATTTACGGCAGGCTGAGTGGCGAGGCTGTGGTGTCCACGTGGCGCTGGGGGGGGTGTGATTCAA  
 15 TGGGCTCGGCTTCCCTGGCCGTGCTGGCGCGCTCACGGGCTTGGGTGGACGCCGACCTCTAGCAGGTGGC  
 TATTCTCCCTTGGAAAGAGAGCCCTCACCATGCTAGGTGTTCCCTGGGTAGGGAGCCTGGTGTGGCAACC  
 CCGGGACCTTAAGCTTATTGTTAAAACATTCTGGGCTGGCTCCGGTGTGCTAAATGGGAAAAGACATCC  
 CACCTCAGCAGAGTTACTGAGAGGCTGAAACGGGGTGTGGCTGACTGGTGTGATCTAGTCATT  
 20 CAGGAAGTCAGTGAGACCAGGTACATGGGGGCTCAGGCAGTGGTGAAGATGAGGTACACGGGGGCTCAGG  
 CAGGGTACATGGGGGCTCAGGCAGTGGTGAAGATGAGGTACACGGGGGCTCAGG  
 GAGGCCAGGTACATGGGGGCTCAGGCAGTGGTGAAGATGAGGTACACGGGGGCTCAGG  
 GAGGCCAGGTACATGGGGGCTCAGGCAGTGGTGAAGATGAGGTACACGGGGGCTCAGG  
 25 GAGGCCAGGTACATGGGGGCTCAGGCAGTGGTGAAGATGAGGTACACGGGGGCTCAGG  
 ACAGGGGCTCTGATCACACCCACATATGAGCACATGTGACATGTGCTTTCATGGT  
 AGCCAGGTCTGTGACACCTGC  
 CCCAAAGTCCCAGGAAGCTGAGAGGCCAAAGATGGAGGCTGACAGGGCTGGCGGGCT  
 CTTTGGGAGGGCGAGGGCGAGAGGATCCCTGAGCCAGGAGTTAAGAC  
 30 CAGCAACACAAACAAAGAAGACTGACAAATGCAGTTCTGGAAAGAAACATTAGTAG  
 GAACTTAACTACACAGGAGGGCTGGCTCAGAAGGGATGCCAGGACGGT  
 GATCACTTGAGCCAGGAGGTGGAGCTGAGTGCACACTGTACTGCAG  
 GGGCATCTCAACAACAAACAAAGAAGACTGACAAATGCAGTTCTGGAAAGAAACATTAGTAG  
 35 GAACTTAACTACACAGGAGGGCTGGCTCAGAAGGGATGCCAGGACGGT  
 GATCACTTGAGCTGGTGTGACCTGTGCCCCATGAGGACATGCA  
 ATCCGTGTGTGACCTGTGCCCCATGAGGAAACCCATGCA  
 ATGCA  
 40 GGGCACACCCACGAGCACCGTCTGATTAGGAGGCTTCTCTGACGCTGTCCGCCATCCTCTCAG

**Intron 14 (SEQ ID NO 18)**

GTATGTGCAGGTGCCCTGGCCTCAGTGGCAGCAGTGCCTGCCGTGGTTAGTGTGTCAGGAGACTGAGTGAATCTGG  
 CTTAGGAAGTCTTACCCCTTTCGCATCAGGAAGTGGTTAACCAACCAACTGTCAGGCTCGTCTGCCCGCCCTCTCGT  
 GGGGTGAGCAGAGCACCTGATGGAAGGGACAGGGAGCTGTCTGGAGCTGCCATCCTCCACCTTGCTCTGCC  
 GCGCTGGGGGCTGGTCTCTCTGTTGCCCATGGTGGATTGGGGGCCCTGCCCTCCTGTTGCCCTGTGGTGG  
 GATTGGGCTGTCTCCGTCATGGCACTAGGGCCCTGTGCAAACCCAGGCCAGGGCTTAGGAGGAGGCCAGGCCAG  
 40 GCTACCCACCCCTCTCAGGAGCACGGCTGTGATTAGGAGGCTTCTCTGACGCTGTCCGCCATCCTCTCAG

5' TCCCTGCCCCCTGGACACTTGTCCAGCATCAGGGAGGTTCTGATCCGTCTGAAATTCAAGCCATTCGAACCTGGGT  
 CCTGAGCTTAACAGCTTCACTTCTGTTCTCTGTGTTGTGGAATTCAACCTGGAGAAGCCGAAGAAAACATTCTG  
 TCGTGACTIONCTGCGGTGCTGGTGGAGCAGCCAGAGATGGAGGCCACCCGAGACCGTGGGTGAGCTTTCCG  
 GTGTCTCCTGGGAGGGAGCTGGCTGGCCTGTGACTCCTCAGCCTCTGTTTCCCCAG

5

**Intron 15 (SEQ ID NO 19)**

5' GCAAGTGTGGTGGAGGCCAGTGCAGGGCCCCACCTGCCAGGGTCATCCTGAACGCCCTGTGTGGCGAGCAGCCTC  
 AGATGCTGAAGTCAGCAGCCCCCGGCCAGCCCTGGAGCCACGCTGGCAGCCATGTGATTAAACG  
 CTGGTGTCCCCAGGGCACGGCCTGGCAGGGTCCCCAACTCTGAAACCCCTGCTTCCATCTCAGGGCGATGGCTC  
 10' CCACGCTTGGGAGGCCTCTGACCCCTGACCTGTGTCCTCTCACAGCCTTCCCTGGCTGTGCCCTGAGCTCTGGGT  
 CCTGAGCAAGTCTCTCCCGCCCGCCCTCAGCTACTGGCTGCTGTGTCCTGCCCTGGGAGGGTGTGAGCTCC  
 TCCCTCACTGAGGTTCCCACAGCCAGGGCACAGGGTGCAGGCCACACGTCTAGGAGG  
 TTGGAGGATGCCACCTCTGGCCTCTGGAAACGGAGTGTGATTGGCCCGCAG

15

**3'-untranskribierter Bereich (SEQ ID NO 20)**

5' ATCTCATGTTGAATCTAATGTGCACTGCAAGACACCACTGTATGCAATTACAGAAGCCTGTGAGTGAACGGGTGGT  
 GGTCAGTGGGGCCCATGGCTGGCTGTCATTACGGAAAGTCTATGAGTGAATGGGTTGTGGTCAGTGGGGCCATG  
 GCCTGGCTGGCCTGGAGGTTCTGATGCTGTGAGGAGGGGAGGAGGGTAGGGATAGACAGTGGAGCCCCCA  
 CCCTGGAAGACATAACAGTAAGTCCAGGGCCGAAGGGCAGCAGGGATGCTGGGGCCAGCTGGGGGGATGATG  
 20' GAGGGCCTGCCAGGGTGGCAGGGATGATGGGGCCCCAGCTGGGGTGGCAGGGGTGATGGGGGGCTGGTCTGGTGG  
 CGGGGAAGATGGGAAGCTGGCTGGGCCCTCCTCCCTGCTCCACCTGAGCCGTGGATCCGGATGTGCTTCCCT  
 GGTGCACATCCTCTGGCCATCAGCTTCACTGGAGGTGGGGCAGGGCATGACACCATCCTGTATAAAATCAGGATT  
 CCTCCCTGCAAGCCCCAACTCAGGTTAAAGTCACATTCCGCTCTGGCATTCTTAAGAGTAGACCAGGATTCTG  
 ATCTGAAAGGGGGTAGGGTGGGAGCAGTGGAGGGTGGACACAGGGCTTCAGGGTGGGCTGGTATGGTCTCTC  
 25' ATCCTCTTATCATCTCCAGTCTCATCTCATCCTCTTATCATCTCCAGTCTCATCTGTCTCCCTTATCTCCAGT  
 CTCATCTGTCATCCTCTTACCATCTCCAGTCTCATCTCTTATCTCTTATCTCTAGTCTCATCCAGACTTACCTCCA  
 GGGGGTGGCCAGGCTCCAGTGGAGCTGGACATACGTCCTCTCAGGCAGAAGGAACGGAAAGGATTGCAAGAACAG  
 GAGGGGGGGCTCAGAGGGACGAGTCTGGGGTGAAGAACAGCCCCCTCTCAGAAGTTGGCTGGCCACACGAAACCG  
 AGGGCCCTGCGTAGTGGCTCCAGAGCCTCCAGCAGGTCCTGGTGGGGCTTATGGTATGGCCGGTCTACTGAGTG  
 30' CACCTGGACAGGGCTCTGGTTGAGTGCAGCCGGAGCTGCGCTGGTGTGGGGGGCTTATGGCACTGGATATG  
 GCGTCATTTATTGCTGCTGCTTCAGAGAATGCTGAGTGCAGGAGCCTAATGTTGATGGGGCCAAGTCCACAGACTG  
 TGTCGAAATGCACTCTGGCTGGAGCCCCGTATAGGAGCTGTGAGGAAGGAGGGCTTGGCAGCCGGCTGGGG  
 GCGCTTCTGCCCTGCAAACCTGGAGGGAGGGCCCGGGCGCCGAGCCTCAAGTGAGAGGTTGGACAGAAC  
 AGGGGGGACTTCCAGAGCAGAGGCCGCTGTCAGGCACACCTGGTTGAATCACAGACCAACAGGTCAGGCCATT  
 35' GTTCAGCTACCATCTTCAAAAGCTCAGATTCTGGTTCTCCGGTGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG  
 TATATTTTTGCTAAAGTATTAGACCTTAAAGGTTAGGGCTTGTATGGCTTAACTCAACTAACGACCTACTTAT  
 TTGCTGTTTATTATTATTATTAGAGATGGTGTACTCTGTCACCCAGGTTGTAGTGCAGTGGCAC  
 AGTCATGGCTGCTGTAGCCGCAAACCCCCAGGCTCAAGTGATCCCGGCCCTCAGCTTCCAGAGTGCTGGGATTACAG  
 GTGTGAGCCACTGCCCTGGCAGCTTTAAAACCACTATGTAAGGTCAAGGTCCAGGTCAGGTCAGGCTTCCACACCTGTCATCC  
 40' CAGTAGTTGGGAAGCCGAGGCAGAAGGATTGTCTGAGGCCAGGAGTTGAGACCAGCATGGTAACATAGGGAGACCC

ATCTCTACAAAAAATGCAAAAGTTATCCGGCGTGGGGTCCAGCATCTGTAGTCCCAGCTGCTGGGAGGCTGAGTGGG  
AGGATCGCTTGAGCCCGGGAGGTATGGCTCAGTGAGCTGTGATTGATTGACATCGACTCCAGCCTGGCAACAGAGTGA  
GACCCCTGTCTCAAAAAAAAAAGAAGGAGAAGGAGAAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG  
GAAGAAGGAAGAAGAAGGAGAAGGAGGCTGCTAGGTGCTAGGTAGACTGTCAAATCTCAGAGCAAATGAAAATAACA  
5 AAGTTTAAAGGGAAAGAAAAACCCAGCTTTGGACTTCCTTAGGCCGAACCTCATCTCAAGCAGCTTCCCTCACA  
GACAACCGGTGTATGGAGCGAGTGAGTTCAAAGCAGAAAGGGAGGAGAACGCAGGCAAGGGTGGAGGCTGTGGGTGACACCA  
GCCAGGACCCCTGAAAGGGAGTGGTTCTGCCTCAGCCCCACGCCCTGCCGTCCTGCACCTGCTGTAACCGTC  
GATGTTGGTGCCAGGTGCCACCTGGAAAGGATGCTGTGCAGGGGGCTGCCAAACTTGGTGGGTTTCAGAAGCCCCAG  
10 GCACTTGTGGCAGGCACAATTACAGCCCCCTCCCAAAGATGCCACGTCTCTCTGCCAAGCTGTGAATGTGTCAACCG  
CAAGGCAGAGGCTGGTGAAGGCTGCAGGTGGAATCACGGCTGCCAGTCAGCCGATCTTAAGGTATCTGGATTATCTGG  
TGGGCCTGATATGGCCACAAGGGTCCCTAGAAGTGAGAGAGGGAGGCAGGGGAGAGTCAGAGAGGGGACGTGAGAAGGAC  
CACTGGCCACTGCTGGCTTGAGATGGAGGAGGGGGTCCCAGCCAAGGAATGGGGCAGCCGCTCCATGCTGGAAAAGC  
AAGCAATCCTCCCGGTCTGAGGGCACACGGCCCTGCCACGCCCTGCCGATTCAGGCCAGTGGGACCTGTTCAGCTTTC  
CGGCCTCCAGAGCTGTAAGATGATGCCCTGTGTTCAAGCCACTAAGCTGCAGTGATTGTCACAGCAGCAAATGGAATAG  
15 CAGTACAGGGAAATGAATAACAGGGACAGTTCTCAGAGTGACTCTCAGCCCCACCCCTGGG

Die Charakterisierung der Exons zeigte interessanterweise, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen, funktionell wichtigen hTC-Protein-Domänen auf separaten Exons angeordnet sind. Das Telomerase-  
5 charakteristische T-Motiv befindet sich auf Exon 3. Die für die katalytische Funktion der Telomerase wichtigen RT (Reverse-Transkriptase)-Motive 1-7 liegen auf folgenden Exons: RT Motiv 1 und 2 auf Exon 4, RT Motiv 4 auf Exon 9, RT Motiv 5 auf Exon 10, RT Motiv 6 und 7 auf Exon 11. RT Motiv 3 liegt verteilt auf Exon 5 und 6 vor (s. Fig. 8).

10

Die Aufklärung der Exon-Intron-Struktur des hTC-Gens zeigt auch, daß die in unserer Patentanmeldung PCT/EP/98/03469 beschriebenen vier Deletions- bzw. Insertions-Varianten der hTC-cDNA ebenso wie drei weitere, in der Literatur (Kilian et al., 1997) beschriebene hTC-Insertions-Varianten höchstwahrscheinlich alternative Splice-Produkte darstellen. Wie in Fig. 8 gezeigt, lassen sich die Splice Varianten in 15 zwei Gruppen einteilen: Deletionsvarianten und Insertionsvarianten.

Den hTC-Varianten der Deletionsgruppe fehlen spezifische Sequenzabschnitte. Die 20 36 bp in frame Deletion in Variante DEL1 resultiert höchstwahrscheinlich aus der Benutzung einer alternativen 3'-Splice Akzeptorsequenz in Exon 6, wodurch ein Teil des RT Motivs 3 verlorengeht. In Variante DEL2 werden die normalen 5'-Splice Donor- und 3'-Splice-Akzeptor Sequenzen von Intron 6, 7 und 8 nicht benutzt. Stattdessen wird Exon 6 direkt an Exon 9 fusioniert, wodurch eine Verschiebung des 25 offenen Leserahmens entsteht und in Exon 10 ein Stopcodon auftritt. Variante Del3 stellt eine Kombination aus Variante 1 und 2 dar.

Die Gruppe der Insertions-Varianten zeichnet sich durch die Insertion von 30 Intronsequenzen aus, die zu vorzeitigen Translationsstop führen. Anstelle der normalerweise benutzten 5'-Splice Donorsequenz von Intron 5 wird eine alternative, 3'-lokalierte Splicestelle in Variante INS1 benutzt, wodurch eine Insertion der ersten 38 bp aus Intron 4 zwischen Exon 4 und Exon 5 entsteht. Ebenso resultiert die

Insertion eines Intron 11-Sequenzbereichs in Variante INS2 aus der Benutzung einer alternativen 5'-Splice Donorsequenz in Intron 11. Da diese Variante in der Literatur (Kilian et al., 1997) nur unzureichend beschrieben wurde, lässt sich die genaue alternative 5'-Splice Donorsequenz dieser Variante nicht bestimmen. Die Insertion 5 von Intron 14 Sequenzen zwischen Exon 14 und Exon 15 in Variante INS3 entsteht durch die Benutzung von einer alternativen 3'-Splice Akzeptorsequenz, wodurch der 3'-Teil von Intron 14 nicht gesplitt wird.

Die in unserer Patentanmeldung PCT/EP/98/03469 beschriebene hTC-Variante INS4 10 (Variante 4) zeichnet sich durch den Ersatz von Exon 15 und dem 5'-Teilbereich von Exon 16 durch die ersten 600 bp des Introns 14 aus. Diese Variante ist auf den Gebrauch einer alternativer interner 5'-Splice Donorsequenz in Intron 14 und einer alternativen 3'-Splice Akzeptorsequenz in Exon 16 zurückzuführen, woraus ein 15 veränderter C-Terminus resultiert.

15 Die *in vivo*-Generation wahrscheinlich nicht-funktioneller hTC-Proteinvarianten, die mit der Funktion des vollständigen hTC-Proteins interferieren könnten, stellt zusätzlich zur Transkriptionsregulation einen möglichen Mechanismus dar, um die hTC-Proteinfunktion zu kontrollieren. Bis heute ist die Funktion der hTC-Splicevarianten nicht bekannt. Obwohl die meisten dieser Varianten vermutlich für 20 Proteine ohne Reverse-Transkriptase-Aktivität kodieren, könnten sie dennoch eine entscheidende Rolle als transdominant-negative Telomerase-Regulatoren spielen, indem sie z.B. um die Interaktion mit wichtigen Bindungspartnern kompetieren.

25 Die Suche nach möglichen Transkriptionsfaktorbindungstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpacket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen für Transkriptionsfaktoren in der Nukleotidsequenz von Intron 2 identifiziert, die in der Tab. 2 aufgelistet sind. 30 Darüberhinaus wurde im Intron 1 eine Sp1-Bindungsstelle (Pos. 43) und im 5'-

untranslatiertem Bereich eine c-Myc-Bindungsstelle (cDNA-Position 29-34, vergl. Fig. 6) gefunden.

#### Beispiel 6

5

Um den oder die Startpunkt(e) der hTC-Transkription in HL 60 Zellen zu ermitteln, wurde das 5'-Ende der hTC-mRNA durch Primer-Extension-Analyse bestimmt.

Es wurden 2 µg PolyA<sup>+</sup>-RNA aus HL-60-Zellen für 10 min bei 65°C denaturiert. Zur 10 Primeranlagerung wurden 1 µl RNasin (30-40 U/ml) und 0,3-1 pmol radioaktiv markierter Primer (5'GTTAAGTTGAGCTTACACTGGTTCTC 3'; 2,5-8x10<sup>5</sup> cpm) zugegeben und für 30 min bei 37°C in einem Gesamtvolumen von 20 µl inkubiert. Nach Zugabe von 10 µl 5xReverse Transkriptase-Puffer (Fa. Gibco-BRL), 15 2 µl 10 mM dNTPs, 2 µl RNasin (s.o.), 5µl 0,1 M DTT (Fa. Gibco-BRL) 2 µl ThermoScript RT (15 U/µl; Fa. Gibco-BRL) und 9 µl DEPC-behandeltes Wasser erfolgte die Primer-Verlängerung in einem Gesamtvolumen für 1 h bei 58°C. Die 20 Reaktion wurde durch 4 µl 0,5 M EDTA, pH 8,0, gestoppt und die RNA nach Zugabe von 1 µl RNaseA (10 mg/ml) für 30 min bei 37°C abgebaut. Hierauf wurden 2,5 µg gescherte Kalbsthymus-DNA und 100 µl TE addiert und einmal mit 150µl 25 Phenol/Cloroform (1:1) extrahiert. Die DNA wurde unter Zusatz von 15 µl 3 M Na-Acetat und 450 µl Ethanol für 45 min bei -70°C gefällt und anschließend für 15 min bei 14000 Upm abzentrifugiert. Das Präzipitat wurde einmal mit 70 %igem Ethanol gewaschen, luftgetrocknet und in 8 µl Sequenzierungs-Stoplösung gelöst. Nach 5 min Denaturierung bei 80°C wurden die Proben auf ein 6 %iges Polyacrylamidgel aufgetragen und elektrophoretisch (Ausubel et al., 1987) aufgetrennt (Fig. 5).

30

Hierbei wurde eine Haupt-Transkriptionsstartstelle identifiziert, die 1767 bp 5' vom ATG-Startcodon der hTC-cDNA Sequenz lokalisiert ist (Nukleotidposition 3346 in Fig. 4). Die Nukleotidsequenz um diesen Haupttranskriptionsstart (TTA<sub>n</sub>TTGT) repräsentiert darüberhinaus ein Initiator-Element (Inr), das in 6 von 7 Nukleotiden

mit dem Konsensusmotiv (PyPyA<sub>n</sub>Na/tPyPy) (Smale, 1997) eines Initiator-Elementes übereinstimmt.

5 In unmittelbarer Nähe des experimentell identifizierten Haupt-Transkriptionsstartes konnte keine eindeutige TATA-Box identifiziert werden, so daß der hTC-Promoter wahrscheinlich in die Familie der TATA-losen Promotoren (Smale, 1997) einzuordnen ist. Allerdings wurde durch Bioinformatik Analyse eine potentielle TATA-Box von Nukleotidposition 1306 bis 1311 (Fig. 4) gefunden. Die zusätzlich um den Haupt-Transkriptionsstart beobachteten Neben-Transkriptionsstarts wurden auch bei 10 anderen TATA-losen Promotoren beschrieben (Geng and Johnson, 1993), wie z.B. in den stark regulierten Promotoren einiger Zellzyklusgene (Wick *et al.*, 1995).

#### Beispiel 7

15 Zusätzlich zu dem in Beispiel 6 beschriebenen, in HL60 Zellen identifizierten Startpunkt des hTC Transkriptes, wurde ein weiterer Transkriptionsstartbereich in HL60 Zellen identifiziert. Anhand von RT-PCR-Analysen wurde die Region des Transkriptionsstarts des hTC-Gens in HL60 Zellen auf die bp -60 bis -105 eingegrenzt.

20 Unter Einsatz von 0,4 µg Poly A-RNA aus HL60 Zellen (Clontech) und dem genspezifischen Primer GSP13 (5'-CCTCCAAAGAGGTGGCTTCTCGGC-3', cDNA-Position 920-897) wurde hierfür die cDNA mit Hilfe des „First Strand cDNA-Synthesis Kit“ (Clontech) nach Angaben der Hersteller synthetisiert. In einem 25 Endvolumen von 50 µl wurden 1 µl cDNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer F (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 10 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 36 PCR-Zyklen an, in denen die DNA zunächst für 30 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer

angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Insgesamt wurden sechs verschiedene 5'-PCR Primer (Primer HTRT5B: 5'-CGCAGCCACTACCGCGAGGTGC-3', cDNA-Position 105 bis 126; Primer CSS: 5'-CTGCGTCCTGCTGCGCACGTGGGAAGC-3', 5'-flankierende Region -49 bis -23; Primer PRO-TEST1: 5'-CTCGCGCGAGTTTCAGGCAG-3', 5'-flankierende Region -74 bis -52; Primer PRO-TEST2: 5'-CCAGCCCCTCCCCTTCCTTCC-3', 5'-flankierende Region -112 bis -91; Primer PRO-TEST4: 5'-CCAGCTCCGCCTCCCGCGC-3', 5'-flankierende Region -191 - -171; Primer RP-3A: 5'-CTAGGCCGATTGACCTCTCTCC-3', 5'-flankierende Region -427 bis -405) mit dem 3'-PCR Primer C5Rück (5'-GTCCCAGGGCACGCACACCAG-3', cDNA-Position 245 bis 225) kombiniert. Als Kontrolle wurde zusätzlich zu den Oligo-dT- und GSP13-geprimten cDNAs auch genomische DNA für die PCR eingesetzt. Wie in Fig. 9 gezeigt, wurde nur mit den Primerkombinationen HTRT5B-C5Rück, C5S-C5Rück und PRO-TEST1-C5Rück ein PCR-Produkt erhalten, was darauf hinweist, daß der Startpunkt der hTC-Transkription in der Region zwischen bp-60 und bp-105 liegt.

#### Beispiel 8

20

In der ca. 11,2 kb isolierten 5'-flankierenden Region des hTC-Gens befinden sich mehrere extrem GC-reiche Bereiche, sog. CpG Islands. Ein CpG Islands mit einem GC-Gehalt von > 70 % reicht von bp - 1214 bis in Intron 2. Zwei weitere GC-reiche Bereiche mit einem GC-Gehalt von > 60 % reichen von bp -3872 bis bp -3113 bzw. bp -5363 bis bp -3941. Die Lage der CpG Islands ist in der Fig. 11 graphisch dargestellt.

30

Die Suche nach möglichen Transkriptionsfaktorbindungstellen wurde mit dem „Find Pattern“-Algorithmus aus dem „GCG Sequenz Analysis“ Programmpaket der „Genetics Computer Group“ (Madison, USA) durchgeführt. Dadurch wurden verschiedene potentielle Bindungsstellen in der Region bis -900 bp upstream vom

Translations-Startcodon ATG indentifiziert: fünf Sp1-Bindungsstellen, eine c-Myc-Bindungsstelle, eine CCAC-Box (Fig. 10). Zusätzlich wurden eine CCAAT-Box und eine zweite c-Myc-Bindungsstelle an den Positionen -1788 bzw. -3995 der 5'-flankierenden Region gefunden.

5

### Beispiel 9

Um die Aktivität des hTC-Promotors zu analysieren, wurden durch PCR-Amplifikation vier verschieden lange hTC-Promotorsequenzabschnitte generiert und 5' vor 10 das Reportergen Luziferase in den Vektor pGL2 der Fa. Promega kloniert. Als DNA-Quelle für die PCR-Amplifikation wurde das aus dem Phagenklon P12 subklonierte, 15 8,5 kb große SacI-Fragment gewählt. In einem Endvolumen von 50 µl wurden 35 ng dieser DNA mit 10 pmol dNTP-Mix versetzt und in 1xPCR-Reaktionspuffer (PCR-Optimizer Kit der Fa. InVitrogen) und einem Unit Platinum-Taq-DNA Polymerase (Fa. Gibco/BRL) eine PCR-Reaktion durchgeführt. Als Primer wurden jeweils 20 20 pmol der nachfolgend definierten 5'- und 3'-Primer zugefügt. Die PCR wurde in 3 Schritten durchgeführt. An eine zweiminütige Denaturierung bei 94°C schlossen sich 30 30 PCR-Zyklen an, in denen die DNA zunächst für 45 sec bei 94°C denaturiert wurde und anschließend für 5 min bei 68°C die Primer angelagert und die DNA-Kette verlängert wurde. Zum Abschluß folgte für 10 min eine Kettenverlängerung bei 68°C. Als 3'-PCR-Primer wurde jeweils der Primer PK-3A (5'-GCAAGCTTGACGCAGCGCTGCCTGAAACTCG-3', Position -43 bis -65) gewählt, der einen Sequenzbereich 42 bp upstream vom START-Codon ATG erkennt. Durch Kombination des PK-3A-Primers mit dem 5'-PCR-Primer PK-5B (5'-CCAGATCTCTGGAACACAGAGTGGCAGTTCC-3', Position -4093 bis -4070) 25 wurde ein 4051 bp großes Promotor-Fragment amplifiziert (NPK8). Die Kombination des Primerpaars PK-3A und PK-5C (5'-CCAGATCTGCATGAAGTGTGGGGATTGCAG-3', Position -3120 bis -3096) führte zur Amplifikation eines 3078 bp großen Promotorfragmentes (NPK15). Ein 2068 bp großes Promotorfragment wurde durch die Verwendung der Primer- 30 kombination PK-3A und PK-5D (5'-

GGAGATCTGATCTGGCTTACTGCAGCCTCTG-3', Position -2110 bis -2087) amplifiziert (NPK22). Der Einsatz der Primerkombination PK-3A und PK-5E (5'-GGAGATCTGTCTGGATTCCCTGGGAAGTCCTCA-3', Position -1125 bis -1102) führte schließlich zur Amplifikation eines 1083 bp großen Promotorfragmente 5 (NPK27). Der PK-3A Primer enthält eine HindIII Erkennungssequenz. Die verschiedenen 5'-Primer enthalten eine BglII-Erkennungssequenz.

Die entstandenen PCR-Produkte wurden mit Hilfe des QIA quick spin PCR Purification Kits der Fa. Qiagen nach Angaben der Hersteller aufgereinigt und 10 anschließend mit den Restriktionsenzymen BglII und HindIII verdaut. Mit den gleichen Restriktionsenzymen wurde der pGL2-Promotor-Vektor verdaut und der in diesem Vektor enthaltene SV40-Promotor freigesetzt und abgetrennt. Die PCR-Promotorfragmente wurden in den Vektor ligiert, in kompetente DH5 $\alpha$ -Bakterien der Fa. Gibco/BRL transformiert. Aus transformierten Bakterienklonen wurde DNA für 15 die nachfolgend beschriebenen Promotor-Aktivitäts-Analysen mit Hilfe des Qiagen Plasmid-Kits der Fa. Qiagen isoliert.

#### Beispiel 10

20 Die Aktivität des hTC-Promotors wurde in transienten Transfektionen in eukaryotischen Zellen analysiert.

Alle Arbeiten mit eukaryotischen Zellen erfolgten an einem sterilen Arbeitsplatz. 25 CHO-K1 und HEK 293 Zellen wurden von der American Type Culture collection bezogen.

CHO-K1 Zellen wurden in DMEM Nut Mix F-12 Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 21331-020) mit 0,15 % Streptomycin/Penezillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) gehalten.

HEK 293 Zellen wurden in DMOD Zellkulturmedium (Fa. Gibco-BRL, Bestellnummer: 41965-039) mit 0,15 % Streptomycin/Penizillin, 2 mM Glutamin und 10 % FCS (Fa. Gibco-BRL) kultiviert.

5 CHO-K1 und HEK 293 Zellen wurden in wassergesättigter Atmosphäre bei 37°C unter Begasung mit 5 % CO<sub>2</sub> kultiviert. Bei konfluentem Zellrasen wurde das Medium abgesaugt, die Zellen mit PBS (100 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7,2; 150 mM NaCl) gewaschen und durch Zugabe einer Trypsin-EDTA Lösung (Fa. Gibco-BRL) abgelöst. Das Trypsin wurde durch Mediumzugabe inaktiviert und die Zellzahl mit 10 einer Neubauer-Zählkammer ermittelt, um die Zellen in gewünschter Dichte auszuplattieren.

15 Für die Transfektion wurden pro Well jeweils 2x 10<sup>3</sup> -HEK 293 Zellen in einer 24-well Zellkulturplatte ausplattiert. Nach 3 Stunden wurde das HEK 293 Medium entfernt. Für die Transfektion wurden bis zu 2,5 µg Plasmid-DNA, 1 µg eines CMV β-Gal Plasmidkonstruktes (Fa. Stratagene, Bestellnummer: 200388), 200 µl serumfreies Medium und 10 µl Transfektionsreagenz (DOTAP der Fa. Boehringer Mannheim) für 15 Minuten bei Raumtemperatur inkubiert und anschließend auf die HEK 293 Zellen gleichmäßig aufgetropft. Nach 3 Stunden wurden 1,5 ml Medium 20 hinzugegeben. Nach 20 Stunden wurde das Medium gewechselt. Nach weiteren 24 Stunden wurden die Zellen zur Bestimmung der Luziferase- und der β-Gal-Aktivität geerntet. Dazu wurden die Zellen im Zellkultur-Lysisreagenz (25 mM Tris [pH 7,8] mit H<sub>3</sub>PO<sub>4</sub>; 2 mM CDTA; 2 mM DTT; 10% Glycerol; 1% Triton X-100) für 15 Minuten bei Raumtemperatur lysiert. Zwanzig µl dieses Zellysats wurden mit 100 µl 25 Luziferase-Assaypuffer (20 mM Tricin; 1,07 mM (MgCO<sub>3</sub>)<sub>4</sub> Mg(OH)<sub>2</sub>·5H<sub>2</sub>O; 2,67 mM MgSO<sub>4</sub>; 0,1 mM EDTA; 33,3 mM DTT; 270 µM Coenzym A; 470 µM Luciferin, 530 µM ATP) gemischt und das durch die Luziferase generierte Licht gemessen.

30 Zur Messung der β-Galaktosidaseaktivität wurden gleiche Mengen Zellysat und β-Galaktosidase-Assaypuffer (100 mM Natriumphosphatpuffer pH 7,3; 1 mM MgCl<sub>2</sub>;

50 mM  $\beta$ -Merkaptoethanol; 0,665 mg/ml ONPG) für mindestens 30 Minuten bei 37°C oder bis eine leichte Gelbfärbung auftrat, inkubiert. Die Reaktion wurde durch Zugabe von 100  $\mu$ l 1 M Na<sub>2</sub>CO<sub>3</sub>, gestoppt und die Absorption bei 420 nm bestimmt.

5 Für die Analyse des hTC-Promotors wurden vier verschieden lange hTC-Promotor-sequenzabschnitte 5' vor das Reportergen Luziferase kloniert (vergl. Beispiel 9).

In der Fig. 11 sind die relativen Luziferase Aktivitäten zweier unabhängiger Transfektionen mit den Konstrukten NPK8, NPK15, NPK22 und NPK27 in HEK 10 293 Zellen aufgetragen. Jedes Experiment wurde in Duplikaten durchgeführt. Darüberhinaus wurde die Standardabweichung angegeben. Das Konstrukt NPK 27 zeigt eine 40fach höhere Luziferaseaktivität als die Basalaktivität des promotorlosen Luziferase-Kontrollkonstrutes (pGL2-basic) und eine 2 bis 3fach höhere Aktivität als das SV40 Promotorkontroll-Konstrukt (pGL2PRO). Interessanterweise wurde im 15 Vergleich zu dem Konstrukt NPK27 eine 2 bis 3fach geringere Luziferaseaktivität in mit längeren hTC Promotorkonstrukten (NPK8, NPK15, NPK22) transfizierten Zellen beobachtet. Ähnliche Ergebnisse wurden auch in CHO Zellen beobachtet (Daten nicht gezeigt).

**Literaturverzeichnis**

5                   **Allsopp, R. C., Vazire, H., Patterson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W. und Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci. 89, 10114-10118.**

10                   **Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). Current protocols in molecular biology. Greene Publishing Associates and Wiley-Intersciences, New York.**

15                   **Blasco, M. A., Rizen, M., Greider, C. W. und Hanahan, D. (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. Nature Genetics 12, 200-204.**

20                   **Broccoli, D., Young, J. W. und deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. Proc. Natl. Acad. Sci. 92, 9082-9086.**

25                   **Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G. Greider, C.W. Harley, C. B. und Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 11, 1921-1929.**

30                   **Feng, J., Funk, W. D., Wang, S.-S., Weinrich, S. L., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. und Villeponteau, B. (1995). The RNA component of human telomerase. Science 269, 1236-1241.**

35                   **Geng, Y., and Johnson, L.F. (1993). Lack of an initiator element is responsible for multiple transcriptional initiation sites of the TATA less mouse thymidine synthase promoter. Mol. Cell. Biol 14:4894.**

**Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. Science 249, 1129-1133.**

**Harley, C.B., Futcher, A.B., Greider, C.W., 1990. Telomeres shorten during ageing of human fibroblasts. Nature 345, 458-460.**

Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D.K. und Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. *Nature* 346, 866-868.

5 Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. und Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. *J. Immunol.* 155, 3711-3715.

10 Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C. B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. und Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011-2015.

Latchman, D.S. (1991). Eukaryotic transcription factors. Academic Press Limited, London.

15 Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. und Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* 276: 561-567.

Lundblad, V. und Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell* 57, 633-643.

20 McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. *Genetics* 26, 234-282.

Meyne, J., Ratliff, R. L. und Moyzis, R. K. (1989). Conservation of the human telomere sequence (TTAGGG)<sub>n</sub> among vertebrates. *Proc. Natl. Acad. Sci.* 86, 7049-7053.

25 Olovnikov, A. M. (1973). A theory of marginotomy. *J. Theor. Biol.* 41, 181-190.

Sandell, L. L. und Zakian, V. A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. *Cell* 75, 729-739.

30 Shapiro, M.B., Senapathy, P., 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* 15, 7155-7174.

35 Smale, S.T. and Baltimore, D. (1989). The "initiator" as a transcription control element. *Cell* 57:103-113.

Smale, S.T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochimica et Biophysica Acta* 1351, 73-88.

5       Shay, J. W. (1997). Telomerase and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.

10       Vaziri, H., Dragowska, W., Allsopp, R. C., Thomas, T. E., Harley, C.B. und Lansdorp, P.M. (1994). Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with age. *Proc. Natl. Acad. Sci.* 91, 9857-9860.

15       Wick, M., Härönen, R., Mumberg, D., Bürger, C., Olsen, B.R., Budarf, M.L., Apte, S. S. and Müller, R. (1995). Structure of the human TIMP-3 gene and its cell-cycle-regulated promoter. *Biochemical Journal* 311, 549-554.

15       Zakian, V. A. (1995). Telomeres: Beginning to understand the end. *Science* 270, 1601-1607.

**Patentansprüche**

1. Regulatorische DNA-Sequenzen für das Gen der humanen katalytischen Telomerase-Untereinheit.

5

2. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um Intronsequenzen gemäß SEQ ID NO 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 und/oder 20 oder um regulatorisch wirksame Fragmente dieser Sequenzen handelt.

10

3. DNA-Sequenzen gemäß Anspruch 1, dadurch gekennzeichnet, daß es sich um die 5'-flankierende regulatorische DNA-Sequenz für das Gen der humanen katalytischen Telomerase-Untereinheit gemäß Fig. 10 (SEQ ID NO 3) oder um regulatorisch wirksame Fragmente dieser DNA-Sequenz handelt.

15

4. Rekombinantes Konstrukt, enthaltend eine DNA-Sequenz gemäß einem der Ansprüche 1 bis 3.

20

5. Rekombinantes Konstrukt gemäß Anspruch 4, dadurch gekennzeichnet, daß es weiterhin eine oder mehrere DNA-Sequenzen enthält, die für Polypeptide oder Proteine kodieren.

6. Vektor, enthaltend ein rekombinantes Konstrukt gemäß Anspruch 4 oder 5.

25

7. Verwendung von rekombinanten Konstrukten bzw. Vektoren gemäß einem der Ansprüche 4 bis 6 zur Herstellung von Arzneimitteln.

8. Rekombinante Wirtszellen, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß einem der Ansprüche 4 bis 6.

30

9. Verfahren zur Identifizierung von Substanzen, die die Promotor-, Silencer- oder Enhanceraktivität der humanen katalytischen Telomerase-Untereinheit beeinflussen, das folgende Schritte umfaßt:

5 A. Zugabe einer Kandidatensubstanz zu einer Wirtszelle, enthaltend DNA-Sequenzen gemäß einem der Ansprüche 1 bis 3, funktionell verknüpft mit einem Reportergen,

10 B. Messung des Substanzeffektes auf die Reportergenexpression.

10 10. Verfahren zur Identifizierung von Faktoren, die spezifisch an die DNA gemäß einem der Ansprüche 1 bis 3 oder an Fragmente davon binden, dadurch gekennzeichnet, daß man eine Expressions-cDNA-Bibliothek mit einer DNA-Sequenz gemäß einem der Ansprüche 1 bis 3 oder Teilfragmenten unterschiedlichster Länge als Sonde screenet.

15 11. Transgene Tiere, enthaltend rekombinante Konstrukte bzw. Vektoren gemäß Ansprüchen 4 bis 6.

20 12. Verfahren zur Detektion Telomerase-assozierter Zustände bei einem Patienten, das folgende Schritte umfaßt:

20 25 A. Inkubation eines rekombinanten Konstruktes bzw. Vektors gemäß Ansprüchen 4 bis 6 das bzw. der zusätzlich ein Reportergen enthält mit Körperflüssigkeiten oder zellulären Proben,

B. Detektion der Reportergenaktivität, um einen diagnostischen Wert zu erhalten,

C. Vergleich des diagnostischen Wertes mit Standardwerten für das Reportergenkonstrukt in standardisierten normalen Zellen oder Körperflüssigkeiten des gleichen Typs wie die Testprobe.

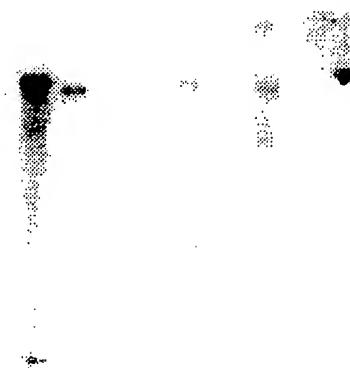
Fig. 1

**A**

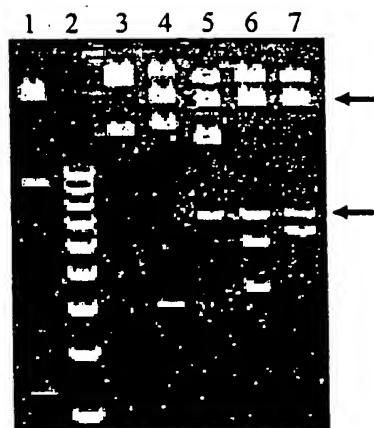
1 2 3 4 5 6 7 8 9 10

**B**

1 2 3 4 5 6 7 8 9 10

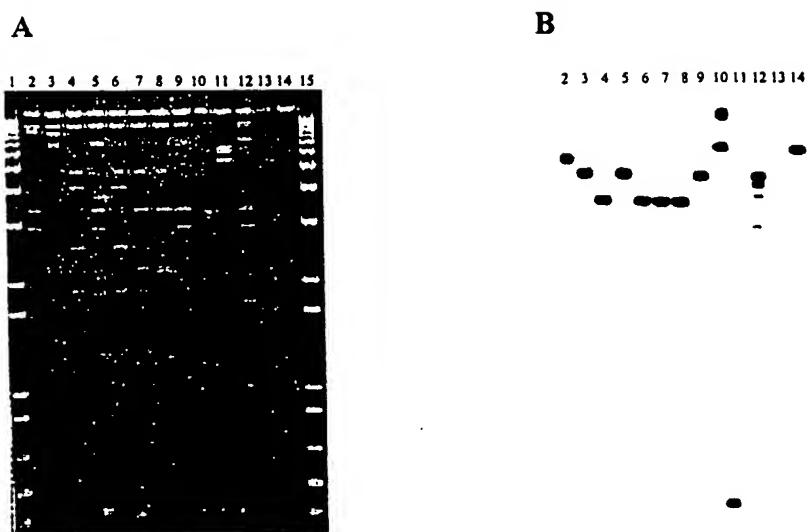
**ERSATZBLATT (REGEL 26)**

**Fig. 2**



**ERSATZBLATT (REGEL 26)**

Fig. 3



**ERSATZBLATT (REGEL 26)**

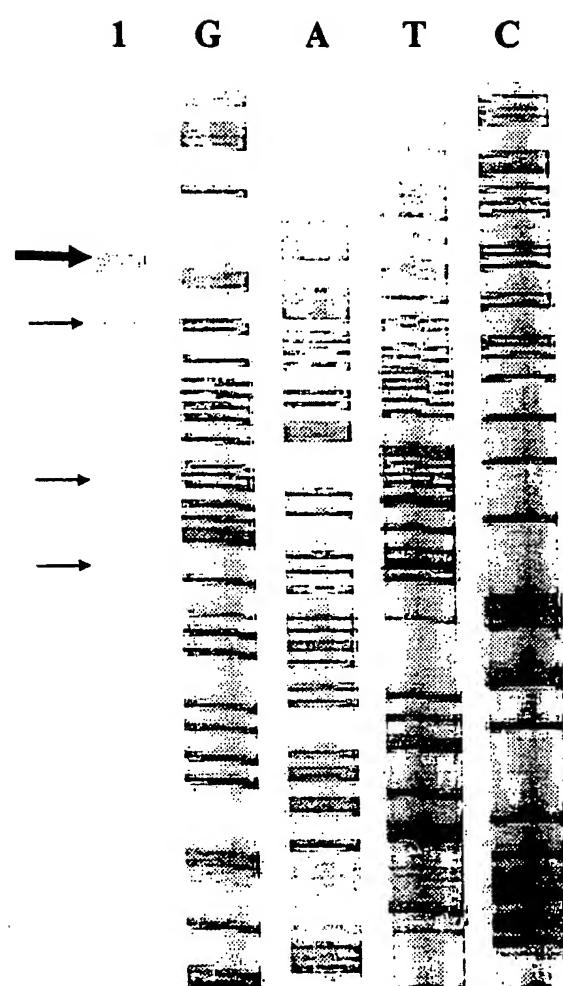
Fig. 4

GAGCTCTGAA CGCTGGAAAC GAACATGACC CTTGCCGCC TGCTTCCCTG GGTGGGTCAA GGGTAATGAA 70  
 GTGGTGTGCA GGAATGGCC ATGTAATTA CACGACTCTG CTGATGGGA CCGTTCCTTC CATCATTATT 140  
 CATCTTCACC CCCAAGGACT GAATGATTCC AGCAACTTCT TCAGGCTGTGA CAAGCCATGA CAAACTCAG 210  
 TACAAACACC ACTCTTTAC TAGGCCACA GACACGGGC CACACCCCTG ATATATAAG AGTCCAGGAG 280  
 AGATGAGGCT GCTTCAGCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTTCC TCTAGACTAG 350  
 TAGACCCCTGG CAGGCACTCC CCCAAATTCT AGGGCCTGGT TGCTGCTTC CGAGGGCCGC ATCTGCCCTG 420  
 GAGACTCAGC CTGGGTGCCC ACACGTGGC CAGGCGCTTC TCCACACCTT CGGCTCCAG GCCTCAGCTT 490  
 CTCCAGCAGC TTCTAAACC CTGGGTGGC CAGTGTCCAG CGCTACTGTC TCAACCTGTC CACTGTGCT 560  
 TGTCAGCC ACCTGACTCG CACGGTCTC CTCACATGG GGTGCTGTC TCCTTCCCA ACACCTCACAT 630  
 GCGTGAAGG GAGGAGATT TGCGCCTCC AGACTGGCTC CTCTGAGGCT GAACCTGGCT CGTGGCCCCC 700  
 GATCAGGTT CCTGGCGTCC GCGTGCACGG TGACCTCCAT TTCCAGGCGC TCCCGCTCTG CTGTCATCTG 770  
 CCGGGGCTG CGGGTGTGTT CTCTGTTCTG TGCTGCTCTT TCCACCTCCA CGTGCAGTGT TCTCTGCCG 840  
 CTAGGCTCTG CGGGTTTTA TACCGCATAGG ACAGGGCGGT CGGGCCAG CGGCTCTTG GGAATGCAA 910  
 CATTGGGTG TGAAGTAGG AGTGCCTGTC CTCACCTAGG TCCACGGCA CAGGCTGGG GATGAGGCC 980  
 CGCGCAGGGA CCCGCCCTTC TCTGCCAGC ACTTTCTGC CCCCCCTCCCT CTGGAACACA GAGTGGCAGT 1050  
 TTCCACAAGC ACTAACATC CTCTTCCCAA AAGACCCAGC ATTGGCACCC CGGACACATT GCCCCACAGC 1120  
 CCTGGGAATT CACGTGACTA CGCACATCTG ATGACACACT CGGTCACAGA CGGACCCCCG CTGTTTATT 1190  
 TTAATAGCTA CAAAGCAGGG AAATCTGC TAAATGTC TTAACAAAC TGTTTAAACAA AACGGTCCA 1260  
 TCCCCACGGT GGACAGTTC CTCACAGTAA GAGGACATG CGGTTTATAA ACCTTCAGG CACATCAAGG 1330  
 GAATTACGCT GAGTCAAAAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTAAAAAAG AAGAATTT 1400  
 ACCCCATGGC AGGGGAGTGG TTAGGGGGGT TAAGGACGGT GGGGGCCGA CGTGGGGGCT ACTGCACGCA 1470  
 CCTTTTACTA AGACCGAGTT CCTGGTTCTG ATGGTATTGG CTCACTTATG GGAGACTAAC CATAAGGGAG 1540  
 TGGGGATGGG CGAACCGGA CGGTGTGCCA TCTTCCCATG CGCGTGTG CTCGGCCAGG ATAATGCTCT 1610  
 AGAGATGCC ACAGTCTGAT TCCCCCAAAC CTGTCAGCAG AACCCGGGG GCCCCAGGG CTTTCAGGT 1680  
 GTGATCTCGG TGAGGACCTT GAGGCTCTGG ATCCCTCGG ACTACCTGCA GGGGGAAAAA GAAATCCAGG 1750  
 GGTCTGGGA AGAGGGGGCC AGGAGGGTCA GAGGGGGGCA CGCTCAGGAC GATGGAGGCA GTCACTGTA 1820  
 GGCTGAAAG GGAGGGAGGG CCTCGAGGCC AGGCGTGCAGA CGCCCTCCAG AACCTGGAAA AACGGGGAA 1890  
 GGGACCTCC CGGAGGCTG CAGGAGAAG CGACGGCTGG CCGTACAGG ACCAGGGCCC ATCGTGACCC 1960  
 TCCGGCTCC CTGCCATAGG AGGGCACTCG CGCTGCCCTT CTAGCATGAA GTGTCGGGG ATTTCAGGAA 2030  
 GCAACAGGAA ACCCATGCA CTCGAATCTA GGATTATTC AAAACAAAGG TTACAGAAA CATCCAAGGA 2100  
 CAGGGCTGAA CTGCTCTCCG GCAACGGCAG GGCAGGCCAG AGTGAATTAA TTACCTTATG TTATTTATT 2170  
 TACTTACTT CTGAGACAGA GTTATGCTCT TGTGCCCCAG GCTGGAGTGC AGCCGATCTGA TCTTGCCTCA 2240  
 CTGCAACCTC CGTCTCTCCG TCTCAAGCAA TTCTCGGCCA CTCAGCTCCC AAGTACCTGG GATTCAAGGC 2310  
 GTGACCAACC ACACCCGGCT AATTGTTATG TTCTAGTAGA GATGGGTTT CACCATGTT GTCAAGCTGA 2380  
 TCTCAAAATC CTGACCTCAG GTGATCCGC CACCTCAGCC TCCCAAAGT CTGGATTAC AGGCATGAGC 2450  
 CACTGCACTT GCGCTATTAA ACCATTTCTG GCTCTCAAGT ACACCCACTG GAAAGGAGTT 2520  
 CATGGAGTTC ATTTCCTCTT TTACTCAGGA GTTACCTCTC TTGATATTT TCTGTAATTC TTCTGAGACT 2590  
 GGGGATACAC CGTCTCTTGA CATAATGACA TTCTCTGTA CCACCTGTTA TCCCATGGGA CCACATGAGC 2660  
 GGGCAGCTGG GAGGCTGCCAG CGTCAAGGTC CCAGTGGGGT TCCCATCTGC CAGTAGAAAC CTGATGAGA 2730  
 ATCAGGGCC AAGTGTGGAC ACTGTCTGA ATCTCAATGT CTCAGTGTG GCTGAAACAT GTAGAAATT 2800  
 AAGTCCATCC CTCTACTCT ACTGGGATTG AGCCCTCTCC CTATCCCCCC CGAGGGCCAG AGGAGTTCC 2870  
 CTCACTCTG TGAGGAGAAG ATGATACTT GTTATTTT CACTGCTGGT ACTGAATCTA CTGTTTCTT 2940  
 TGTGGTTTG TTGTTTTGT TTGAGGCG CGTCTACTG TTGTTCTCA CGCTGGAGGG AGTGCATGG 3010  
 CGCGATCTTG GCTTACTGCA CGCTCTGCC CCCAGGTCA AGTCACTTC CTGCTTCGG CTCCCAATTG 3080  
 GCTGGGATTA CAGGCCACCC CGACCATGCC CAGCTAATTG TTGTTATTT TAGTAGAGAC GGGGGTGGGT 3150  
 GGGGTTCAACG ATGTTGGCCA CGCTGGCTCT GAACCTCTGA CCTCAGATGA TCCACCTGCC TCTGCCCTCC 3220  
 AAAGTGTGG GATTACAGGT GTGAGCCAC ATGCCAGACT CAGAATTTAC TCTGTTAGA AACATCTGG 3290  
 TCTGAGGCTG GAAAGCTCACC CGACTCAAGT TTGTTCTGTT TTAACAGCAA TGATGAAATT TTATTTATTG 3360  
 TGTAGAACAA CTCTGTATG TTTACACTGT GATGACTAAG ACATCATCAG CTTTCMAAG ACACACTAAC 3430  
 TCCACCCATA ATACTGGGT GCTCTCTGG TATCAGCAAT CTTCACTGAA TGCCGGAGG CGTTCCCTCG 3500  
 CCATGCACTT GTGTTAATT ACTCCACAT ATCTCTCTG TTCCATTCTC TCTCTCCCT TTATTTAAAT 3570  
 TGTGTTCTCT ATGTTGGCTT CTCTGAGAG AACCAGTGTG AGCTACAACT TAACCTTGT TGGAAACAAAT 3640  
 TTCTCAAACC GCCCCCTTGC CCTAGTGGCA GAGAACATC ACAACACAG CCCTTAAAAA AGCCTTAGGG 3710  
 ATCACTAAGG GGATTTCTAG AAGAGCGACCC TGTAACTCTA AGTATTTACA AGACGAGGCT AACCTCCAGC 3780  
 GACCGTGACA GCCCAGGGAG GGTGGAGGC CGTTCAAAT GCTACCTCCA TAAATAAAGC AATTTCCCTC 3850  
 GGCAGTTCT GAAAAGTAGGA AAGGTTACAT TAAAGGTGCG TTGTTCTGAG ATTTCACTGT TTGCGGACCT 3920  
 CAGCTACAGC ATCCCTGCCA GGCCTGGGA GACCCAGAAG TTCTCGGCC CCTTGAATCC AACACTGAGC 3990  
 AACCCGGAGT CTGGATTCTC GGGGAGTCT CAGCTGTCTG CGGGTTGTGC CGGGGGCCCA GGTCTGGAGG 4060  
 GGACCACTGG CGGTGTGGCT TCTACTGCTG GGTGGGAAGT CGGGCTCCCT AGCTCTGCCAG TCCGAGGCTT 4130  
 GGACCCAGGT CGCTGGACCC CGAGGCTGCC CTCCACCTG TGCGGGGGGG ATGTCACAG AGTGTGGCCT 4200  
 CATCTGCCAG ACAGACTGCC GGGGCCCTAG GTCAAGGGCCG TTGTTGCTGC TGTTGAGGCGC CGGGTGGCGC 4270  
 GCCACCCAGA CGGCCCTGGCT CCATTTCCCA CCTTCTCTG ACGGGACCC CGGGTGGGT GATTAACAGA 4340  
 TTGGGGATG TGAGGACCC CTCCCCGGCT GAGAACCTGC AAAGAGAAAT GACGGGCCCTG 4410  
 TGTCAAGGAG CCCAAGTGC GGGGAAGTGT TGCAAGGGAGG CACTCCGGGA GGTCCCGCGT GCGCGTCCAG 4480  
 GGACCAATGC GTCCCTGGGT TGTGCCCCAG CGCGCTCTAC CGCGCTCCGT CCTCCCTTC ACGTCCGGCA 4550  
 TTCTGGTGC CGGGAGCCCC AGGCCCCCGC TCCGGACCTG GAGGCAAGCCC TGGGTCTCCG GATCAGGCCA 4620  
 CGGGCCAAAG GTGCGCCCA CGCACCTGTT CCCAGGGCT CCACATCATG GCCCCCTCCCT CGGGTTACCC 4690

Fig. 4 (Fortsetzung)

CACAGCCTAG	GCCGATTCGA	CCTCTCTCCG	CTGGGGCCCT	CGCTGGCGTC	CCTGCACCCCT	GGGAGCGCGA	4760
GGGGCGCGCG	GGGGGGAAAG	GGCGGCCAG	ACCCCCGGGT	CCGGCCGGAG	CAGCTGGCT	GTGGGGCCA	4830
GGCGGGGCTC	CCAGTGGATT	GGCGGGCACA	GACGCCAGG	ACCGCGCTCC	CCACGTGGCG	GACGGACTGG	4900
GGACCCGGC	ACCCGTCTCG	CCCCCTTCACC	TTCCAGCTCC	GCCTCTCCG	CGCGGACCCC	GCCCCGTCCCC	4970
GACCCCTCCC	GGGTCCCCGG	CCCAGCCCC	TCCGGGGCCCT	CCCAGCCCC	CCCCCTTCCTT	TCCGGGGCCC	5040
CGCCCTCTCC	TGGGGCGCGG	AGTTTCAGGC	ACCGCTGGGT	CCTGCTGCGC	ACGTGGGAAG	CCCTGGCCCC	5110
GGCCACCCCCC	<u>GGGATG</u>						5126

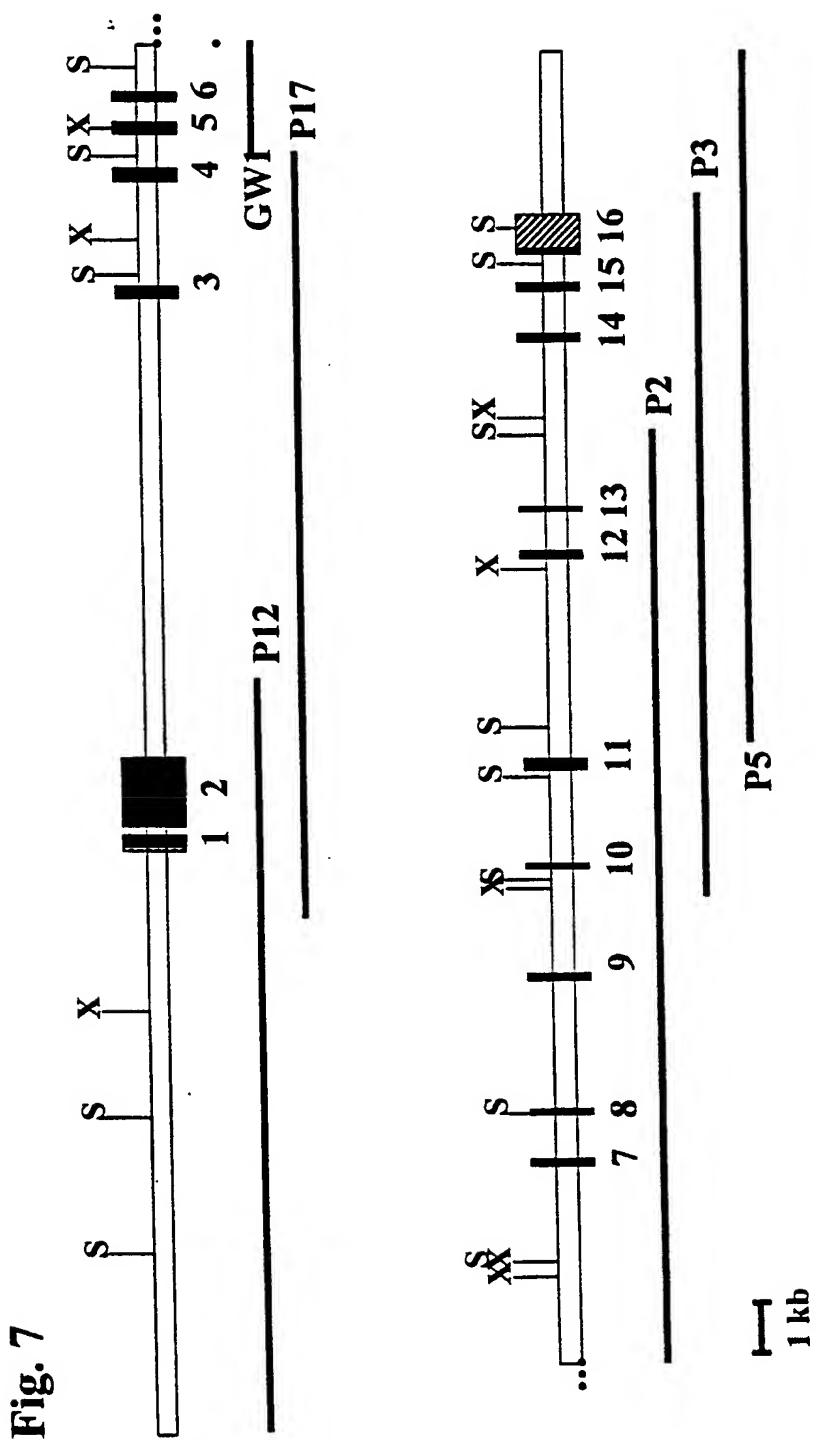
Fig. 5



ERSATZBLATT (REGEL 26)

Fig. 6

GTTCAGGCA GCGCTGCGTC CTGCTGCGCA CGTGGGAAGC CCTGGCCCG GCCACCCCG CGATGCCCG 70  
 CGCTCCCGC TGCGGAGCCG TCGCTCCCT GCTGCGCAGC CACTACCGCG AGGTGCTGCC GCTGGCCACG 140  
 TTCGTGCGGC GCCTGGGGCC CCAAGGGCTGG CGGCTGGTGC AGCGCGGGGA CCCGGCGGCT TTCCCGGCC 210  
 TGGTGGCCCA GTGCTGGTG TCGCTGCCCT GGGACGCGACG GCGCCCGCCG GCGCCCGCCCT CCTTCGCGCA 280  
 GGTGCTCTGC CTGAGGGAGC TGGTGGCCCG AGTGTGCGAG AGCGCTGCGG AGCGCGGCCG AAGAACGCTG 350  
 CTGCCCCCTCG GCCTCGCGCT GCTGGACGGG GCGCGGGGG GCGCCCGCCGA GGCCTTCACC ACCACGCGTC 420  
 GCAGCTACCT GCGAACACAG CTGACCGACG CACTGCGGGG GAGCGGGGGG TGGGGGCTGC TGCTGCGCCG 490  
 CGTGGCGAC GACGTGCTGG TTCACCTGCT GGCACGCTGC CGCGCTTTC TGCTGGTGGC TCCCAGCTGC 560  
 GCCTACCAAGG TGTCGGGGCC CCGCGCTGAC CAGCTCGCG CGCCACTCA CGCCCGGCCG CGCCACACAG 630  
 CTAGTGGACC CGCAAGGGCGT CTGGGATGCC AACGGGCGCT GACCATAGC GTCAGGGAGG CGGGGGTCCC 700  
 CCTGGGCGCTG CGAGCCCCGG CGCGAGGGAG GCGCGGGGG AGTGCAGGCC GAGACTGCGC GTTGCCCAAG 770  
 AGCCCCAGGC GTGGGCGCTGC CGCTGAGGGC GAGCGGACCC CGGTGGGCA GGGGGCTCTGG GCGCCACCCGG 840  
 GCAGGACGGG TGAGCCGAGT GACCGTGGTT TCTGTGTTG GTCACCTGCC AGACCCCGCG AAGAACCCAC 910  
 CTCTTGGAG GGTGCGCTCT CTGGCACCGG CGACTCCCAAC CCATCGCTGG CGCGCCAGCA CCACCGGGCG 980  
 CCCCCATCCA CATCGCGGC ACCACGTCCT GGGACACGCC CTGTGCCCC GGTGTAACGCC GAGACCAAGC 1050  
 ACTTCTCTA CTCCCTCAGGC GACAGGAGC AGCTCGGGCC CTCCCTCTCA CTCACTCTC TGAGCCCCAG 1120  
 CCTGACTGGC GCTCGGAGGC TCGTGGAGAC CATTTCTG GGTTCAGGC CTCGGATGCC AGGGACTCCC 1190  
 CGCAGGTTGC CGCCGCTGCC CGACGCTAC TGGCAAATGC GGGCCCTGTT TCTGGAGCTG CTTGGGAAACC 1260  
 ACGCCAGTG CCCCTACGGG GTGCTCTCA AGACCGACTG CCCGCTGCGA GTCGCGGTCA CCCACGCGAC 1330  
 CGGTGCTGTG GCGCCGGAGA AGCCCCAGGG CTCTGTGGC GCGCCCGAGG AGGAGCACAC AGACCCCCGT 1400  
 CGCCTGGTGC AGCTGCTCG CGACGACAGC AGCCCCCTGGC AGTGTACGG CTTCTGCGG GCCCTGCTGC 1470  
 GCGCCCTGGT GCGCCCGAGG CTCTGGGGCT CGAGGCACAA CGAACCCCGC TTCTCAGGA ACACCRAGAA 1540  
 GTTCATCTCC CTGGGGAAGC ATGCCAAGCT CTCGCTGCCAG GAGCTGRCGT GGAAGATGAG CGTGCAGGAC 1610  
 TGCCTTGGC TGCGCAGGAG CCCAGGGGTT GGCTGTGTC CGGCGCCAGA GCACCGCTCG CGTGAGGAGA 1680  
 TCCCTGGCAA GTTCCTGCAC TGGCTGATGA GTGTGTAAGT CGTGCAGCTG CTCAGGTCTT TCTTTATGT 1750  
 CACGGAGACC AGCTTCAAA AGAACAGGCCT TTTTTCTAC CGAAGAGCTG TCTGGAGCCTA GTTGCAAGAC 1820  
 ATTGGAAATCA GACAGCACTT GAAGAGGGTG CAGCTGGGG AGCTGTCGGA ACCRAGGTC AGGCACCATC 1890  
 GGGAAAGCCAG GCGCCGCGCTG CTGACGTCCA GACTCCGTT CATCCCCAAG CCTGACGGGC TCGGGCGAT 1960  
 TGTAACATG GACTACGTCG TGGAGCCAG AACGTTCCGC AGAGAAAGA GGGCGGAGCG TCTCACCTCG 2030  
 AGGGTGAAGG CACTGTTCA CGTGCCTCAAC TACGAGCGGG CGCGGGCCCC CGGCTCTCTG GGGCGCTCTG 2100  
 TGCTGGGCTG GGACGATATC CAACAGGGCT CGCGCACCTT CGTGTGCGT GTGGGGGCC AGGACCCGCC 2170  
 GCGCTGAGCTG TACTTTGTCA AGGTGGATGT GACGGGGCGG TAGCACACCA TCCCCCAGGA CAGGCTCAGC 2240  
 GAGGTGATCG CCAGCATCAT CAAACCCAG AACACGTAAC CGCGCGCTCG GTATGCCGTG GTCCAGAAGG 2310  
 CGGCCCATGG GCACTCGCC AAGGCCCTCA AGAGCCAGT CTCTACCTTG ACAGACCTCC ACCCGTACAT 2380  
 GCGACAGTTC GTGGCTCAC TGCAGGAGAC CAGCGCGCTG AGGGATGCC CGTGTATCGA CGAGAGCTCC 2450  
 TCCCTGAATG AGGCCAGCAG TGGCTCTTC GACGTCTCAT TACGCTTCAAC CGCGTGGCAGA 2520  
 TCAGGGCCAA GTCTACGTC CAGTCCCGAG CGATCCCCA GGGCTCCATC CTCTCACCG TGCTCTGCG 2590  
 CCTGTGCTAC GCGGACATGG AGAACAAAGCT GTTGGGGGG ATTCCGGGG ACGGGCTGCT CCGCGTTTG 2660  
 GTGGATGATT TCTTGTGTT GTGACCTTCAC CTCACCCACG CGAAAACCTT CCTCAGGACCT CTGTTCCGAG 2730  
 GTGTCCTCTGA TGTGGCTGCC GTGGTAAGAC AGTGGTQAAC TTCCCTGTAG AAGACGAGGC 2800  
 CCTGGGTGGC AGCGCTTTTG TTCAAGATGCC GCGCCACGGCT CTATCCCCCT GTGCGGCGCT GTGCTGGAT 2870  
 ACCCGGACCC TCGAGGTGCA GAGCGACTAC TCCAGCTATG CGCGCACCTC CATCAGAGCC AGTCTCACCT 2940  
 TCAACCGGG CTTCAAGGCT GGGAGGAACA TGCCTGGCAA ACTCTTGGG GTCTTGGCC TGAGTGTC 3010  
 CAGGCTGTTT CTGGATTTCG AGGTGAACAG CGCCAGCGACG GTGTGACCAAC ACATCTACAA GATCCTCTG 3080  
 CTGCAAGGGCT ACAGGTTCTA CGCAATGTTG CTGACGCTTC CATTTCATCA GCAAGTTGG AAGAACCCCCA 3150  
 CATTTTCTC GCGCGTCATC TCTGACAGCG CCTCCCTCTG CTACTCCATC CTGAAAGCCA AGAACCGAGG 3220  
 GATGTCCTGC GGGGCCAAGG GCGCCGGGG CCCTCTGCC TCCGAGGGCC TGCAGTGGCT GTGCCACCAA 3290  
 GCAATTCTGC TCAAGCTGAC CGACACCGT GTCACCTACG TGCCACTCC GGGGTCACTC AGGACAGGCC 3360  
 AGACCGACGT GAGTCGGAGC CTCCCGGGGA CGACGCTACG TGCCTGGAG GCGCGAGCCA ACCCGGCACT 3430  
 GCGCTCAGAC TTCAAGACCA TCTGGACTG ATGGCCACCC GCGCACAGCC AGGGCGAGAG CAGACACCG 3500  
 CAGCCCCGTC ACCGGGGGCT CTACGCTCCA GGGAGGGAGG GCGCCCGCCAC ACCCAGGGCC CGACCGCTGG 3570  
 GAGTCTGAGG CCTGAGTGA GGTGGGGCC AGGCGCTGAT GTCCGGCTGA AGGCTGAGTG TCCGGCTGAG 3640  
 GCCTGAGGGA GTGTCCAGCC AAGGGCTGAG TGTCCAGCAC ACCTGCCGTC TTCACCTCCC CACAGGCTGG 3710  
 CCCTCGGCTC CACCCCGAGGG CCAGCTTTTC CTACCCAGGA GCGCCGCTTC CACTCCCCAC ATAGGAATAG 3780  
 TCCATCCCCA GATTCGCGCAT TGTTCACCCC TCGCCCTGCG CTCCCTTGTCC TTCCACCCCC ACCATCCAGG 3850  
 TGGAGACCCCT GAGAAGGACC CTGGGAGCTC TGGGAATTGG GAGTGCACCAA AGGTGTCCCC TGACACAGG 3920  
 CGAGGACCCCT GCACCTGGAT GGGGGTCCCT GTGGGTCAAA TTGGGGGAG GTGCTGTGGG AGTAAATAC 3990  
 TGAATATATG AGTTTTTCAG TTTTGAAGA AAAAAAAA AAAAAAAA AA 4042



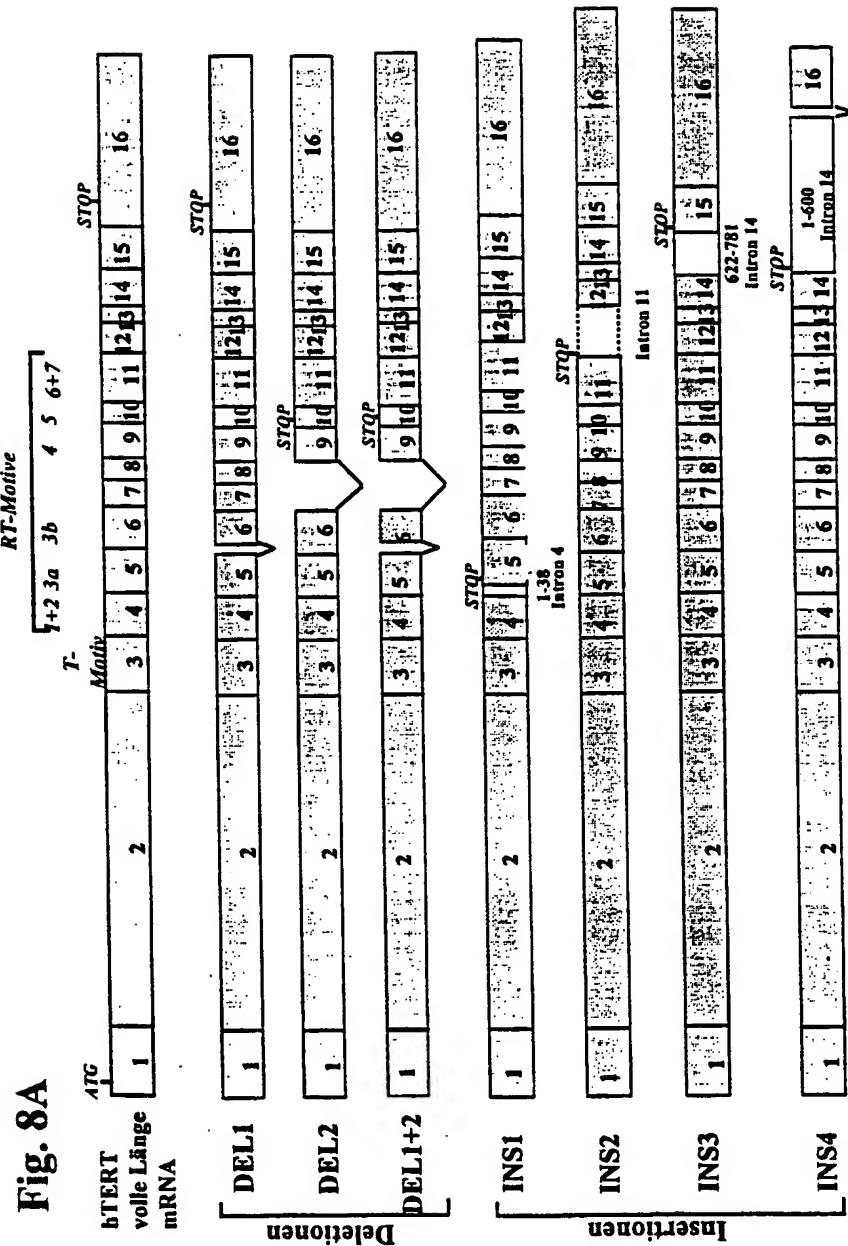
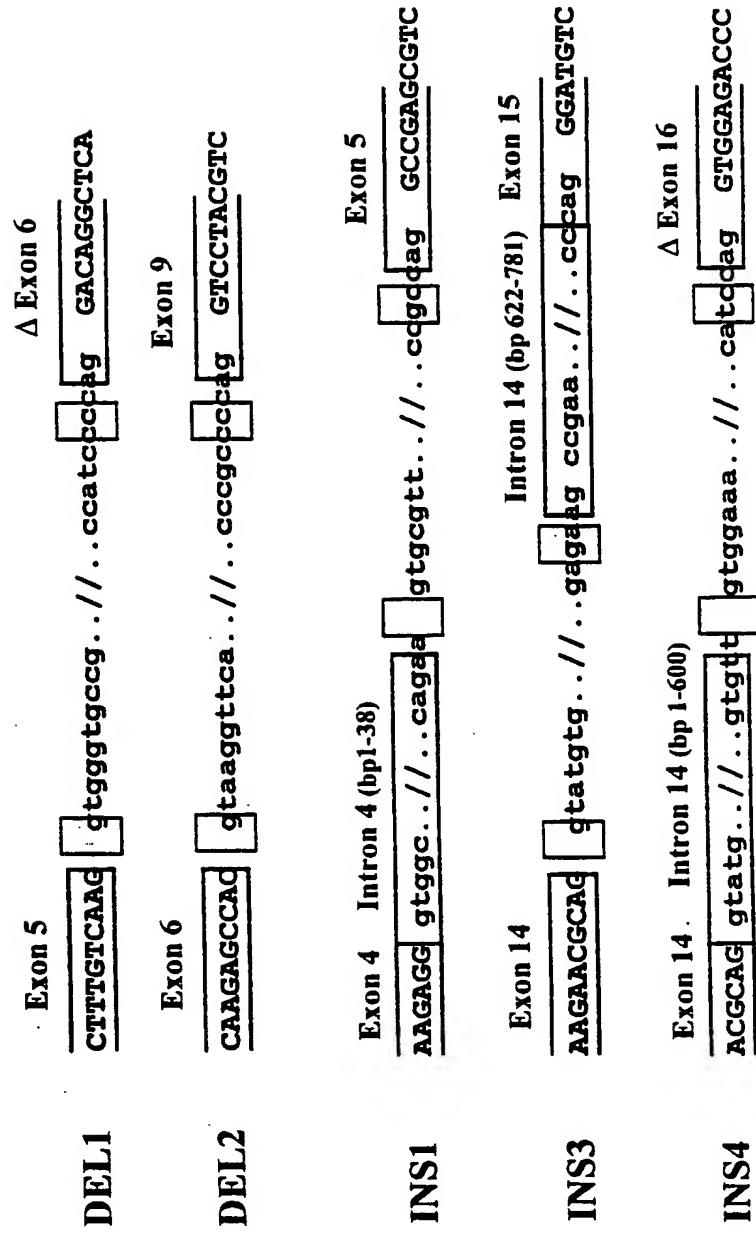


Fig. 8B



ERSATZBLATT (REGEL 26)

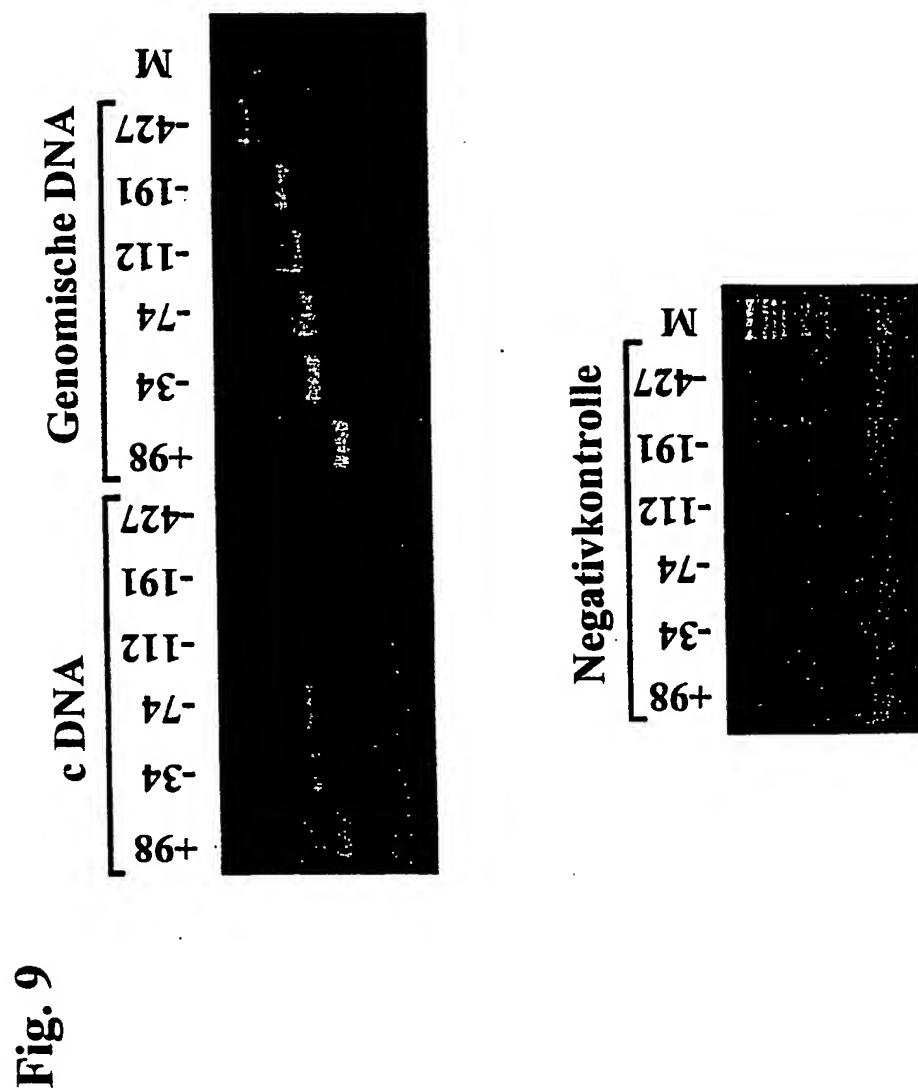


Fig. 9

Fig. 10

ACTTGAGCCC AAGAGTTCAA GGCTACGGTG AGCCATGATT GCAACACCAC AGCCGACGCT TGGTACAGA -11204  
 ATGAGACCT GTCTCAAAAA AAAAAAAA AATTGAAATA ATATAAAGCA TCTTCTCTGG CCACAGTGG -11134  
 ACAAACCG AAATCAACAA CAAGAGGAAT TTGAAAATC ATACAAACAC ATGAAAATTA AACATATAC -11064  
 TTCTGAATGA CCACTGAGTC AATGAAGAAA TTTAAAAGGA AATTGAAAAA TTTATTTAAG CAAATGATAA -10994  
 CGGAAACATA ACCTCTCAA ACCCACGGTA TACAGCAAA GCAGTGCTAA GAAGGAAGTT TATAGCTATA -10924  
 AGCAGCTACA TCAAAAAAGT AGAAAAGCA GGCAGCTGG CTCATGCCTG TAATCCCAGC ACTTTGGAG -10854  
 GCCAAGGCGG GCAGATCGCC TGAGGTCAAG AGTCAGAGAC CAGCTGACC AACACAGAG AACCTTGTAG -10784  
 CTACTAAAAA TACAAATTA GCTGGGCAATG GTGGCACATG CTCGTAATC CAGCTACTCG GGAGGCTGAG -10714  
 GCAGGATAAC CGCTTGAACC CAGGGAGGTGG AGGTTGCGGT GAGCGGGAT TGCAGCCATT GACTCCAGCC -10644  
 TGGGTAACAA GAGTGAACCT CTGTCCTAAG AAAAAAAA AAGTGAAGAA ACTTAAAAT ACAACCTAAT -10574  
 GATGCACCTT AAAGAACTAG AAAAGCAAGA GCAAACCTAA CCTAAATATG GTAAAAGAAA AGAAATAATA -10504  
 AAGATCAGAG CAGAAATAAA TGAAGATGAA AGATAACATC ACACAAAGATC AACAAATTA AAAGTGGTT -10434  
 TTTGAAAAG ATAACAAATC TTGACAAACC TTGACCAAGA CTAAGAAAAGG AAGGAAAGAG ACCTAAATAA -10364  
 ATAAAGTCAG AGATGAAAAA AGAGACATTA CAACTGATAC CACAGAAAAT CAAAGGATCA CTAGAGGCTA -10294  
 CTATGAGCAA CTGTACACTA ATAAATTGAA AAACCTAGAA AAAATAGATA AATTCTCTAGA TGCACTACAAC -10224  
 CTACCAAGAT TGAACCATGAA AGAAATCCAA AGCCAAACAA GACCAATAAC AATAATGGGA TAAAGCCAT -10154  
 ATAAAGAACT CTCTTAGCAA AGAGAAGGCC AGGACCAAT GGCTTCCCTG CTGGATTTA CCAATCATTT -10084  
 AAAGAGAAAT GAATTCCAAT CCTACTCAA CTATCTGAATC AAATAGAGGA AAGAATACTT CCAAACTCAT -10014  
 TCTACATGGC CAGTATTACG CTGATTCCAA ACCAGACAA AAACACATCA AAAACAAACA AACAAAAAAA -9944  
 CAGAAAGAAA GAAAATCACA GCCCAATATC CCTGATGAAT ACTGATACAA AAATCCTCAA CAAAACACTA -9874  
 GCARACCAA TAAACACAA CCTTCGAAAG ATCATTCACT GTGATCAAGT GGGATTATT CCAGGGATGG -9804  
 AAGGATGGTT CAACATATGC AAATCAATCA ATGTGATACA TCATCCCAAC AAAATGAAGT ACACAAACTA -9734  
 TATGATTATT TCACTTATG CAGAAAGAAC ATTTGATATAA ATTCGACCT CTCATGATA AAAACCTCA -9664  
 AAAAACAGG TATACAAGAA ACATACAGGC CAGGCACAGT GGCTCACCC TGCAGCTGGC GCACTCTGG -9594  
 AGGCCAAGGT GGGATGATTG CTGGGGCCCA GGAGTTGAG ACTAGCTGG GCAACAAAT GAGACCTGGT -9524  
 CTACAAAAAA CTTTTTAAAT AAATTAGCCA GGCATGATGG CATATGCCTG TAGTCCCGC TAGTCTGGAG -9454  
 GCTGAGGTGG GAGAATCACT TAAGCCTAGG AGGTCGAGGC TGCACTGAGC CATGAACATG TCACTGTACT -9384  
 CCAGCTAGA CAACAGAACAA AGACCCCACT GAATAAGAAG AAGGAGAAGG AGAAGGGAGA AGGGAGGGAG -9314  
 AAGGGAGGAG GAGGAGAAGG AGGAGGTGGA GGAGAGTGG AGGGAGGAGA GAGGAAGAAG -9244  
 AAGAAACATA TTTCACATA ATAAAGGCC TATATGACAG ACCGAGGTAG TATTATGAGG AAAAATCTGAA -9174  
 AGCCTTCCTC CTAAGATCTG GAAAATGACA AGGGCCACT TTCAACACTG TGATTCAACA TAGTACTAGA -9104  
 AGTCCCTAGCT AGAGCAATCA GATAAGAGAA AGAAATAAAAA GGCATCCAA CTGAAAGGA AGAAGTCAAA -9034  
 TTATCCTGTT TGCAGATGAT ATGATCTTAT GCTGAAATTG GGTACAGCG GATACAAAAAAT CAAATGACAA AAATCTCTATA TTCCAAACAGC -8894  
 AAACAATCTG AAAAAGAACAA CAAAGGAGCA GTGAAAGATC TCTACATGAA AAACATCTAA GCTAGGAAATT AACCAAAGAA -8824  
 AGATGAGACA TGGACCAAGAG GAACAGAAATA TTTTGACAA AGGTAATCTG CTCACCCACT TGATTCAACA TAGTACTAGA -8754  
 AGATATTCCA TGTTCATAGA TTGAAAGAAT AAATACTGTT AAAATGTCCA TACTACCCAA AGCAATTTC -8684  
 AAATTCAATG CAATCCCTAT TAAATACTA TTGTCAGAA CCACAAAAGA CCAAGCTAT CTCGACCAAA AGAACACAAA CTGGAAGCAT -8614  
 CACATTACCT GACTCTAAAT TATACCTAAC AGCTATAGTA ACCCAAACAA CATGGTACTG GCATAAAACAC -8474  
 AGATGAGACA TGGACCAAGAG GAACAGAAATA TTTTGACAA AGGTAATCTG CTCACCCACT TGATTCAACA TAGTACTAGA -8404  
 CTGGATATCC ATATGCAAA TAACAATACTA GGAGTAAAGG CTTAAATCTA AACACCTCAA CTTTGCAACT ACTAAAAGAA AACACCCGAG AAACCTCTCA -8194  
 GGACATTGGA GTGGGCAAG ACTCTCTGAG AAATGGGATC ATATCAAGTT AAAAAGCTTC CCACAGCAAA AGAACACAGC -8124  
 CCACAGAAATG GGAGAATATA TTGCAACT GTCATCTAA CAAAGGATTA ATAACCAATCC ATGCATCTAC AGTGAACCTCA -8054  
 ATTCACTCAA CAAGGAAATTA ATAACCAAGTA TATATAAGGA -7984  
 GCTCAACTA CTCTATAAGA AAAACACCTA CATTGCAACT ATACAAACAA AGAACACAAA ATCTCAACTC CCACATTTAC -7914  
 CATTCTCAA ATAAGTCTAT CAAATGGCA GAGAAATGCA AATAGGAGAG AGGCAATAAC AAATGCACTG CTCACACCA CTGATCATCA -7844  
 ATCATCTCAT CCCAGTTAA ATGGCTTTA TTCAAAGAG -7774  
 AGGCAATAAC AAATGCCAGT GAGGATGTGG ATAAAAGGAA ACCCTTGAC ACTGTTGGTG GGAATGGAA -7704  
 TTGCTACAC TATGGAGAAC AGTTGAAAG TTCTCAAA AACTAAAAAAT AAAGCTACCA TACAGCAATC -7634  
 CCATTGCTAG GTATATACTC CAAAAAAGGG AATCAGTGTCA TCAACAAAGCT ATCTCAACTC CCACATTTAC -7564  
 TGCAGCACTG TTACATGAGC CCAAGGTTG GAAGCACACCT CAGTGTCTC CAACAGACGA ATGGAAAAAG -7494  
 AAAATGTGGT GCACATACAC ATGGAGTAC TACGCGACCA TAAAAGGAA TGAGATCTG TCAGTTGCAA -7424  
 CAGCATGGGG GGCAGCTGGC AGTATGTTAA GTGAAATAAG CCAGGCACAG AAAGACAAAC TTTTCATGTT -7354  
 CTCCCTTACT TGTGGGAGCA AAAATAAAAA CAATTGACAT AGAAATAGAG GAGAATGGTG GTTCTAGAGG -7284  
 GGTGGGGAC AGGGTACTA GAGTCAACAA TAATTATG TATGTTTAA AATAACTAA AGAGTATAAT -7214  
 TGGGTTGTTT GTACACAAA GAAAGGATAA ATGCTTGAAG TGAGACAGATA CCCCCTTAC CTCGATGTGA -7144  
 TTATTACACA TTGTTATGCT GTATCAAAT ATCTCATGTA TGCTATAGAT ATAACCCCTA CTATATAAA -7074  
 AATTAAAATT TTAATGGCCA GGCACGGTGG CTCTATGCTG TAATCCCAGC ACTTTGGAG GCGCAGGGCGG -7004  
 GTGGATCACC TGAGGTCAAG AGTTGAAAC CAGTCTGGCC ACCATGATGA AACCTGTCT CTACTAAAGA -6934  
 TACAAAAAATT AGCCAGGCCT GGTGGCACAT ACCCTGAGTC CCAACTACTC AGGAGGCTGA GACAGGAGA -6864  
 TTGCTTGAAC CTGGGAGGGC GAGGTTGCGAG TGAGCCGAGA TCATGCCACT GCACTGAGC CTGGGTGACA -6794  
 GAGCAAGACT CCATCTCAA AAAAAAAACAA AAAAAGAG ATTAATTTAAG TAATTCTTAT GTACCGTATA -6724  
 AATATATAC TCACTATATT AGAAGTAAA AATTAAACAA ATTATAAAAG GTAATTAAACC ACTTAATCTA -6654  
 AAATAAGAAC AATGTATGTG GGGTTCTAG CTTCTGAAGA AGTAAAGTT ATGGCCACGA TGGCAGAAAT -6584

Fig. 10

GTGAGGGAGGG AACAGTGGAA GTTACTGTTG TTAGACGCTC ATACTCTCTG TAAGTGACTT AATTGAAACC -6514  
 AAAGACAGGG TGGGAGAAGT TAAAGAGGCA TTCTATAAGC CCTAAAACAA CTGCTAATAA TGGTGAAGG -6444  
 TAATCTCTAT TAATTACAA TAATTACAGA TATCTCTAA ATCGAGCTGC AGAATTGGCA CGTCTGATCA -6374  
 CACCGTCCTC TCATTCACGG TGCTTTTTT CTTGTGTGCT TGAGGATTTC CGATTGTGTT TTCGTGTTG -6304  
 GTTAAACTTA ATCTGTATGA ATCTGAAAC GAAAATGGT GGTGATTTC TCCAGAAGAA TTAGAGTACC -6234  
 TGGCAGGAAG CAGGTGGCTC TGTGACCTG AGCCACTTC ATCTTCAAGG GTCTCTGGCC AAGACCCAGG -6164  
 TGCAGGCAAG AGGGCTGATG AGCCGAGGAC AGGAAAGCTC GGATGGAAG GGGCGATGAG AAGCCTGCCT -6094  
 CGTTGGTAG CAGCGATCA AGTGCCTTA TTACGCTT GCAAGGATTG CTCTGGATAC CATCTGAAA -6024  
 AGGCCGCAG CGGGAAATGCA AGGAGTCAGA AGCCTCTGC TCAAACCCAGG GCCAGCAGCT ATGGCGCCA -5954  
 CCCGGCGTG TGCCAGAGGG AGAGGAGTCA AGGCACCTCG AAGTATGGCT TAAATCTTTT TTTCACCTGA -5884  
 AGCAGTGACC AAGGTGATT CTGAGGGAAAG CTTGAGTTAG GTGCTCTCTT TAAAACAGAA AGTCATGAA -5814  
 GCACCCCTCT CAAGGGAAA CCAGACGCC GCTCTGCGT CATTACCTC TTCTCTCTC CCCTCTTTG -5744  
 CCCTCGGGT TTCTGATCGG GACAGAGTGA CCCCCCTGGA GCTTCTCGA GCCCCTGCT AGGACCCCT -5674  
 TGCAAAGGGC TCCACAGACC CCCGCCCTGG AGAGAGGAGT CTGAGGCTGG CTTAATTAACA AACTGGGATG -5604  
 TGGCTGGGGG CGGACAGCGA CGGGGGGATT CAAAGACTTA ATTCCATGAG TAAATTCAAC CTTTCCACAT -5534  
 CCGAATGGAT TTGGATTTTA TCTTAAATT TTCTTAAATT TCATCAAATA ACATTCAGGA CTGCAGAAT -5464  
 CCAAGGCGT ARAACAGGA CTGAGCTATG TTTGCCAAGG TCCAAGGACT TAATAACCAT GTTCAGAGGG -5394  
 ATTTTGCCTC CTAGTACTT TTATTTGGT TTCAAAAGGT GGCTTAGGGT GCAAGGGAAA GTACACGAGG -5324  
 AGAGGCCCTGG CGGGCAGGGC TATGACGACG CAGGGCCAC CGGGGAGAGA GTCCCCGGCC TGGGAGGCTG -5254  
 ACAGCAGGAC CACTGACCGT CCTCCCTGG AGCTGCCACA TTGGGCAACG CGAAGGGCCG CACGCTGCCT -5184  
 GTGACTCAGG ACCCCATACC GGCTCTCTGG GCCCACCCAC ACTAACCCAG GAAGTCACGG AGCTCTGAAC -5114  
 CCGTGGAAAC GAACATGACCC CTTGCCCTGG TGCTCTCTG GGTGGTCAA GGGTAATGAA GTGGTGTGCA -5044  
 GGAATGGCC ATGTAATTA CACGACTCTG CTGATGGGGG CTTCTCTCC CATCATTATT CATCTTCACC -4974  
 CCGAAGGACT GAATGATTCC AGCAACTTCTC TCGGGTGTG CAAGCCATGA CAAACACTCG TACAAACACC -4904  
 ACTCTTTAC TAGGCCACA GAGCACGGSC CACACCCCTG ATATATTAAG AGTCCAGGAG AGATGAGGCT -4834  
 GCTTCAGCC ACCAGGCTGG GGTGACAACA GCGGCTGAAC AGTCTGTCC TCTAGACTAG TAGACCCCTG -4764  
 CAGGCACTCC CCCAGATCTC AGGGCTCTGG TGCTCTCTC CGAGGGCCG ATCTGCCCTG GAGACTCAGC -4694  
 CTGGGGTGCAC ACTGAGGGC CAGCCCTGTC TCCACACCTC CCGCTCCAG GCTCAGCTT CTCCAGCAGC -4624  
 TTCTAAACCC CTGGGTGGGC CGTGTTCAGG CGTACTGTG TCACCTGTCC CACTGTGCT TGCTCAGCG -4554  
 ACGTAGCTCG CACGGTTCTC CCTCACATGG GGTGCTGTG TCCTTCCCCA ACACTCACAT GCGTTGAAGG -4484  
 GAGGAGATTG TGCGCTCTCC AGACTGGCTC CTCTGAGCCT GAACCTGGCT CGTGGGCCCG GATGCGAGGTT -4414  
 CCTGGCGTCC GGCTGCACGC TGACCTCCAT TTCCAGGCGC TCCCCGTCCTC CTGTCATCTG CGGGGCCCTG -4344  
 CCGGTGTGTT CTCTGTTTC TGTGCTCTT TCCACGTCA GCTGGCTGTG TCTCTGGCC CGGGGTCTC -4274  
 GGGGTTTTA TAGGCATAGG ACGGGGGGGT GGTTGGCCAG GGCGCTCTG CGAAATGCAA CATTGGGTG -4204  
 TGAAAGTAGG AGTGCCTGTC CTCACCTAGG TCCACGGGCA CAGGCCCTGG GATGGAGGCC CGGCCAGGG -4134  
 CCCGCCCTC TCTGCCAGC ACTTCTCTGC CCCCCCTCCCT CTGGAACACA GAGTGGCACT TTCCACAAAGC -4064  
 ACTAAGCATC CTCTTCCAA AAGACCCAGC ATTGGCACCC CTGGACATT GCCCCACAGC CCTGGAAATT -3994

*c-Myc*

CACCTGACTA CGCACATCAT GTACACACTC CCGTCCACCG CCGACCCCCC CTGTTTATT TTAATAGCTA -3924  
 CAAAGCAGGG AAATCCCTGC TAAAATGTCC TTTAACAAAC TGTTAAACA AACGGGTCCA TCCGCACGGT -3854  
 GGACAGTCC TCACAGTGAAG GAGGAACATG CGGTGTTAA AGCCTGAGG CATCTCAAGG GAATTACGCT -3784  
 GAGTCAAACAC TGCCACCTCC ATGGGATACG TACGCAACAT GCTCAAAAGG AAAGAATTTC ACCCCATGGC -3714  
 AGGGGAGTGG TTAGGGGGT TTAGGACGGT GGGGGGGCA CTGGGGGGT ACTGCAACGC CCTTTTACTA -3644  
 AAGCCAGTTT CCTGGTTCTG ATGGTATTGG CTCAGTTATG GGAGACTAAC CATAGGGAG TGGGGATGGG -3574  
 GGAACCCGGA GGCTGTGCCA TCTTGGCCAT GCCCAGTGT CTTGGCAGG ATAATGCTCT AGAGATGCC -3504  
 ACGTCTGAT TCCCCAACAC CTGGGACAG AACCCGGCC GCCCCAGGGC CTTTGCAGGT GTGATCTCCG -3434  
 TGAGGACCCCT GAGGTCTGGG ATCTCTGGG ACTACCTGCA GGGCCGGAAA GTAATCCAGG GGTCTGGGA -3364  
 AGAGGGGGCAGGAGGGTCA GAGGGGGGCA GCCTCAGGGC GATGGAGGCA GCTAGCTGA GGCTGAAAAG -3294  
 GGAGGGAGGG CCTCGAGGCC AGGCGCTGAA GCGCCTCCAG AAGCTGGAAA AAGCGGGGAA GGGACCCCTC -3224  
 ACGGAGCTG CAGCAGGAAG GCACGGCTGG CCCTTAGCCC ACCAGGGCCC ATCGTGGACC TCCGGCCTCC -3154  
 GTGCCATAGG AGGGCACTCG CGCTGCCCCCT CTAGCATGAA GTGTTGGGG ATTGGCAGAA GCAACAGGAA -3084  
 ACCCATGCAC TGTGAATCTA GGATTATTTT AAAACAAAGG TTACAGAAA CATCCAAGGA CAGGGCTGAA -3014  
 GTGCCCTCCGG CGCAAGGGCAG AGTGTATTA TTAGCTGTT TTTAGCTT ATTATTTATT TACTTACTTT -2944  
 CTGAGACAGA GTTATGCTCT GTTGGCCAG GCTGGAGTGC AGCGGCATGA TCTTGGCTCA CTGCAACCTC -2874  
 CGTCTCTGGG GTTCAAGCAA TTCTCGTGC TCAGCCTCCC AAGTAGCTGG GATTTCAGGC GTGCACCAAC -2804  
 ACACCCGGT AATTGTTAT TTTAGTAGA GATGGGGTTT CACCATGTG TGCAAGCTGA TCTCAAAATC -2734  
 CTGACCTCGAG GTGATCCGCC CACCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCATGAGC CACTGCACCT -2664  
 GGCTATTAA ACCATTAA AACTCTCTG GGCTCAACGT ACACCCACTG GTAAAGGAGTT CATGGAGTTC -2594  
 AATTCCCCCT TTACTCAGGA GTTACCTCC TTGATATT TCTGTAATT TCTGTAGACT GGGGATACAC -2524  
 CGTCTCTGG TTTGAGCAA GAGGCTGAG GCTTCAGGTC CCAGTGGGGT TGCCATCTGC CAGTAGAAAC CTGATGTAGA ATCAGGGCGC -2384  
 AAGTGTGGAC ACTGTCTGA ATCTCAATGT CTCAAGTGTG GCTGAAACAT GTAGAAATTAA AAGTCCCATCC -2314  
 CTCTTACTCT ACTGGGATTG AGCCCCCTCC CTATCCCCCC CCAGGGCAG AGGAGTCTC CTCACTCTG -2244  
 TGGAGGAAGG ATGATACTT GTTATTTTT CACTGCTGGT ACTGAATCCA CTGTTTCATT TGTGGTTG -2174  
 TTTGTTTTGT TTGAGAGGC GGTTTCACTC TTGTTGCTCA GGCTGGAGGG AGTGCATGG CGGGATCTTG -2104  
 GCTTACTGCA GCCTCTGCCT CCCAGTTCA AGTGAATTCTC CTGCTTCCGC CTCCCATTT GCTGGGATTA -2034  
 CAGGCACCCG CCACCATGCC CAGCTAATT TTTGATTTT TAGTAGAGAC GGGGGGGT GGGGTTCCACC -1964

Fig. 10

ATGTTGGCCA GGCTGGTCTC GAACTTCTGA CCTCAGATGA TCCACCTGCC TCTGCCTCCT AAAGTGTGG -1894  
 GATTACAGGT GTGAGCCACC ATGCCAGCT CAGAATTAC TCTGTTAGA AACATCTGGG TCTGAGGTAG -1824  
*C44T-Box*  
 GAAGCTCACC CCACTCAAGT GTTGTGGTGT TTTAAGCAA TGATAGAAATT TTTTATTGT TGTTAGAAC -1754  
 CTCTTGATGT TTTACACTGT GATGACTAAG ACATCATCAG CTTTCAAAG ACACACTAAC TGCAACCCATA -1684  
 ATACTGGGGT GTCTTCTGGG TATCAGCAAT CTTCATTGAA TGCGGGGAGG CGTTTCTCG CCATGCACAT -1614  
 GGTGTTAATT ACTCCAGCAT AATCTTCTGC TTCCATTCT TCTCTTCCCT CTTTAAAT TGTGTTTCT -1544  
 ATGTTGGCTT CTCTGCAGAG ACCAGTGTAA GCTACAACT TAACTTTGT TGGAACAAAT TTTCCAAAC -1474  
*Spl*  
GCCCTTGC CCTAGTGGCA GAGACAAATC ACAAACACAG CCCCCTAAAAA AGGCTTAGGG ATCACTAAGG -1404  
 GGATTCTAG AAGAGCGACC TGTAACTCTA AGTATTTACA AGACGAGGCT AACCTCCAGC GAGCGTGACA -1334  
 GCCCAGGGAG GGTGCGAGGC CTGTTCAAAAT GCTAGCTCCA TAAATAAAGC AATTCCTCC GGCAAGTTCT -1264  
 GAAAGTAGGA AAGGTTACAT TTAAGGTTGC TTGTTAGC ATTTCACTGT TTGCGGACCT CAGCTACAGC -1194  
 ATCCCTGCAA GGCTCGGGA GACCCAGAAG TTTCTCGCCC CCTTAGATCC AAACTTGAGC AACCCGGAGT -1124  
 CTGGATTCTT GGGAAAGTCCT CAGCTGTCTT GCGGGTTGTGC CGGGGGCCCA GGTCTGGAGG GGACCAAGTGG -1054  
 CCGTGTGGCT TCTACTGCTG GGCTGGAAGT CGGGCCTCTT AGCTCTGCAG TCCGAGGCTT GGAGCCAGGT -984  
 GCCTGGACCC CGAGGCTGCC CTCCACCCCTG TGCGGGCGGG ATGTGACCAAG ATGTTGGCT CATCTGCCAG -914  
 ACAGAGTGC C GGGGCCAGG GTCAAGGCCG TTGTTGGCTGG TGTGAGGCCG CGGTGCGCG GCCAGCAGGA -844  
*CCAC-Box*  
 GCGCCTGGCT CCATTCCCAC CCCTTCTCG ACAGGACCCGC CCCGGTGGGT GATTAACAGA TTTGGGGTGG -774  
 TTTGCTCATG GTGGGGACCC CTGCGCCCT GAGAACCTGC AAAGAGAAAT GACGGGCTG TGTCAAGGAG -704  
 CCCAAGTCGC GGGGAAGTGT TGCAGGGAGG CACTCCGGGA GGTCCCCCGT GCCCCTCCAG GGAGCAATGC -634  
*AP-2*  
 GTCCCTGGGT TCGTCCCCAG CGCGTCTAC GCGCCTCCGT CCTCCCTTC ACGTCCGGCA TTGCGGGTGC -564  
 CGGGAGCCCG ACGGCCCGCG TCCGGACCTG GAGGCAGCCC TGGGTCTCG GATCAGGCCA GCGGGCAAAG -494  
 GGTCCCGCA CGCACCTGTT CCCAGGGCCT CCACATCATG GCCCCTCCCT CGGGTTACCC CACAGCTAG -424  
*Spl*  
 GCCGATTGCA CCTCTCTCCG CTGGGGCCCT CGCTGGCGTC CCTGCACCC GGGAGCGCGA GCGGGCGCG -354  
GGGGGGAAAG CGCGGCCAG ACCCCCCGGGT CGCGCCGGAG CAGCTGCCCT GTCGGGGCCA GGCGGGCTC -284  
 CCAGTGGATT CGCGGGCACA GACCCCCAGG ACCGGCCCTCC CCACGTGCG GAGGGACTGG GGACCCGGGC -214  
 ACCCGTCTG CCCCTTCACC TTCCAGCTCC GCTCTCTCCG CGCGGACCC GCCCCGTCCC GACCCCTCCC -144  
*Spl*  
 GGGTCCCCGG CCCAGCCCCC TCCGGGCCCT CCCAGCCCC CCCCTTCCCT TCCGGGCCCG CGCCCTCTCC -74  
*c-Myc*  
 TCGCGGCCCG AGTTTCAGGC AGCGCTGCGT CCTGCTGCC ACGTGGAAG CCCTGGCCCC GGCCACCCCC -4  
 GCGATG

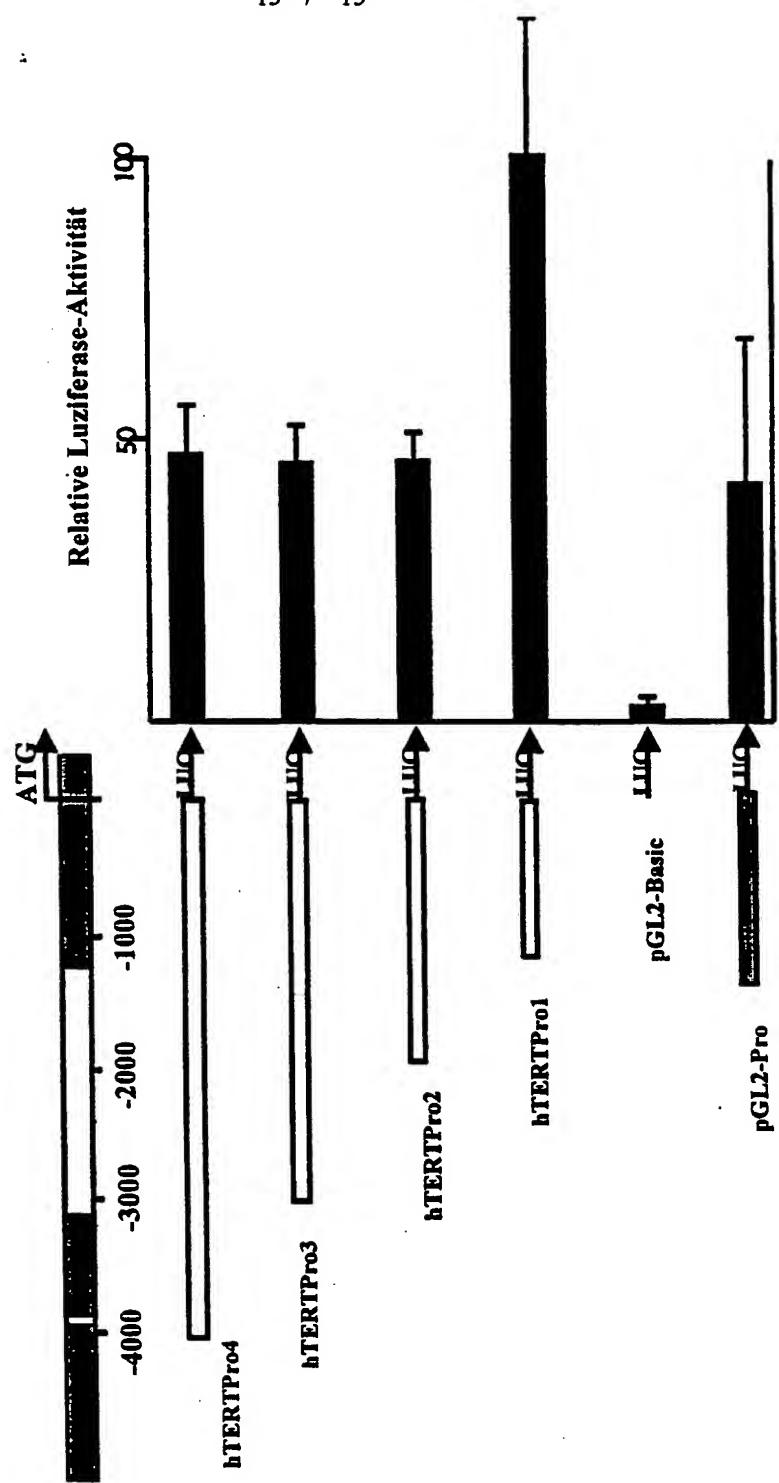


Fig.: 11

## SEQUENZPROTOKOLL

<110> Bayer AG

5 <120> Regulatorische DNA-Sequenzen aus der 5i-Region vom Gen  
der humanen katalytischen Telomerase-Untereinheit und  
deren diagnostische und therapeutische Verwendung

10 <130> LeA32805-Ausland

15 <140>

<141>

20 <160> 20

<170> PatentIn Vers. 2.0

25 <210> 1

<211> 5126

<212> DNA

<213> Homo sapiens

30 <400> 1

gagctctgaa ccgtggaaac gaacatgacc cttgcctgcc tgcttccctg ggtgggtcaa 60  
gggtatgaa gtgggtgtca ggaaatggcc atgtaaaattt cacgacttcc ctgatggga 120  
ccgttcccttccatcatttcatcatttccacc cccaaaggact gaatgattcc agcaacttct 180  
tcgggtgtca caagccatgaa caaaacttag tacaaacacc actctttac taggcccaca 240  
gagcacgggc cacaccccttccatataatggc agtccaggag agatgaggct gcttccagcc 300  
accaggcttgg ggtgacaaca ggggctgaaac agtctgtcc tctagactag tagacccttgg 360  
caggcactcc cccaaatttccatcatttccacc cccaaaggact gaatgattcc agcaacttct 420  
gagactcagc ctgggggtgcc acactgaggcc cagccctgtc tccacaccccttccggcc 480  
gcctcagggtt cttccagcagc ttccctaaacc ctgggttggc cgtgttccagc cgtaactgtc 540  
tcacctgttcc cactgtgtcc tggcttcaggc agcttagetcc caccgggttcc tccatcatgg 600  
gggtgttgc tccctcccaacatcaccatc ggttggaaagg gaggagattc tgccgccttcc 660  
agacttggcc tcttgagcccttccacc cccaaaggact gaatgattcc agcaacttct 720  
ggctgcacgc tgacccatccatcatttccacc cccaaaggact gaatgattcc agcaacttct 780  
ccgggtgttcc tttctgttttccatcatggc tccacgttcc gttggctgttcc tcttgccttcc 840  
ctagggttcc ggggttttccatcatggc tccacgttcc gttggctgttcc tcttgccttcc 900  
ggaaatgcaaa cattttgggttccatcatggc tggaaatggtagg agtgcctgtcc tccacccatggc tccacggcc 960  
caggccttggg gatggggcccttccacc cccaaaggact gaatgattcc agcaacttct 1020  
cccccttcccttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1080  
aagaccggcatttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1140  
cgcacatcatgttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1200  
caaaaggcagggttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1260  
tccgcacgggttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1320  
catcttcagggttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1380  
gctcaaaaaggaaatggtagg tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1440  
ggggggggca gctggggggcttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1500  
atggatattgg tccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1560  
ggctgttgc tccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1620  
acgttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1680  
gtgtatctccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1740  
gtaatccagggttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1800  
gatggaggca gtcagttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1860  
gcccctccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1920  
gcacggctgtcc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 1980  
aggccactcttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 2040  
acccatcatgttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 2100  
cagggttgc tccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 2160  
ttatcttcatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 2220  
agccggcatgttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 2280  
tcagcttccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 2340  
ttttatgtatcatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 2400  
gtgtatccatcatggc tccacccatggc tggaaatggtagg agtgcctgtcc tccacccatggc 2460

```

<210> 2
<211> 4042
<212> DNA
50  <213> Homo sapiens

<400> 2
gttcaggcca ggcgtgcgtc ctgctgcgtc cgtgggaagc cctggcccg gccaccccg 60
cgatgcccg cgctccccgc tgccgagccg tgccgtcccc gctgcgcgcg cactaccccg 120
55  aggtgctgcc gctggccacg ttctgtgcggc gcctggggcc ccagggctgg cggctggtgc 180
agccggggga ccccgccgtc ttccggcgcg tggtgcgcga tgccgtggcg tgccgtgc 240
gggacgcacg ggcggccccc cccgtccgcga ggtgtccctgc ctgaaaggagc 300
55  tggtggcccg agtgcgtgcag aggctgtgcgc agccggccgc gaaagaaatgc ctggccctcg 360
gcttcggcgt gctggacccgg gccegggggg gcccccccgaa ggccttcacc accagcgtgc 420
60  gcagctaccc gcccaacacg gtgacccacg cactgcgggg gaggccggcg tgggggctgc 480
tgctgcgcg cgtggggcgcac gacgtgtctgg ttccactgtgc ggcacgcgtgc ggcgtcttcg 540
tgctgggtgc tcccaactgc gcctacccagg tggtggggcc gccgtgtac cagtcggggc 600
ctggccactca ggccggggccccc ccggccacacg ctatggacc cccgaaggcgt ctggatgcg 660
65  aacggggctcg gaaccatagg gtcaggaggcc cccgggtccc ctggggctcg ccagcccccgg 720
qtqcgaaqqaq qcqcqggggc agtgcacggc gaactgtgc gttgccccaaag agggcccgaggc 780

```





5 gaaaggataa atgcttgaag gtgacagata ccccattac cctgatgtga ttattacaca 4140  
ttgtatgcct gtatcaaataat atctcatgtat tgctatagat ataaacccta ctatattaaa 4200  
aattaaaaat ttaatggcca ggcacgggtgg ctcatgtcc taatcccacg accttgggag 4260  
gcccggggcg ggtagatcctc tgaggtcagg agtttgaac cagtcgtcc accatgtatg 4320  
aaccctgtct ctactaaca taaaaaaaatc agccaggcgt ggtggccat acctgtatgc 4380  
ccaactactc aggaggctga gacaggagaa ttcttgcacg ctgggaggcgc gaggttgcag 4440  
tgagccgaga tcatgtccact gcactgcgc ctgggtgaca gagcaagact ccatctcaaa 4500  
acaaaaaacaa aaaaaagaag attaaaatttgc taatttttgc ttaatccatgc aatataact 4560  
ctactatata agaagttaaa aaaaaacaa attataaaatg gtaattaaacc acttaatct 4620  
aaataagaac aatgtatgc gggtttctat cttcttgcgaa agttaaaaatg atggccacg 4680  
tggcagaaat gtgaggaggg aacagtggaa gttactgttg tttagacgctc atactctc 4740  
taagtactt aattttaaacc aaaaacggcgc tggtggatgg taaagggca ttatataacg 4800  
cctaaaaacaa ctgctaataa tggtgaaagg taatctctat taattaccaaa taattacaga 4860  
tatctctaaa atcgagctgc agaattggca cgtctgtatca caccgtcc tcattcacgg 4920  
tgcttttttctt ctgtgtgtcg tggagattt ctgtgtgttg ttctgtgttg gttaaactt 4980  
atctgtatga atctcttgc aaaaaatgtt ggtgtatcc tccagaaagaa tttagatacc 5040  
tggcaggaaat caggtggcgc tggtggaccc tggacacttcc atcttcagg gtctctggc 5100  
aagaccggcagg tgcaggcgc aggtctgtatg accggaggac agggaaatgc ggtatggaa 5160  
gggcgtatgag aagctgtccct cggtggtag cagcgcgtatg agtgccttta ttacgtttt 5220  
gcaaaagattg ctctggatcatc catctggaaa aggccggcag cgggaatgcg aggagtccg 5280  
agccctctgc tcaaaccctcg gccagcgtt atggccggcc cccgggctgtg tgccagggg 5340  
agaggatgttcc aggcacccctcg aagtatgtgtttaatctttt ttcacacttgc agcagtgcacc 5400  
aagggttattt ctggggatgg ctggatgttag tgccctttt taaaacagaa atgtatggaa 5460  
gcacccctctt ctggggaaaatcc cccacgcggc gctctgtgggtt cattttacttcc ttcttcctt 5520  
ccctctcttg ccctcgccgtt ctgtgtatcg gacagatgtt gccccgtggaa gcttctccg 5580  
gccccgtgtcg aggaccctctt tgcaaaaggcc tccacagacc cccgccttgg agagaggagt 5640  
ctgagcttgc ttaataaca aactgggtatg tggtgtgggg cggacacgcg cggccggatt 5700  
caaagactta atccatgtatg taaaatccacat ctttccatcatc cggatgtatg ttggatattt 5760  
tcttaatattt tttttttttat tcatcaataatc acatccatggat ctgcagaaat ccaaaacggcgt 5820  
aaaacaggaa ctgagctatg ttttccaaagg tccaaaggact taataaccat gttcagagg 5880  
atttttcgcc ctaagtaactt tttttttttt ttcataaaggat ggccttgggtt gcaaggggaaa 5940  
gtacacgagg agggccctgg gccgcaggcc tatgacacg cgcggccac cggggagaga 6000  
gtccccccggcc tggggggctg acacggggatc cactgtccctt gcttcccttgg agtcgcacca 6060  
ttggggcaacg cgaaggccgcg cactgtgtt gttacttgcgg accccatccatc ggttctccg 6120  
gcccccccccactaaccctcg aactatccatgg agctctgtatc acctgtggaaac ccgtggaaac 6180  
cttgccttgc tgggttttttgg ggtgggtcaaa gggtaatgaa ctgggtgtca gggaaatggcc 6240  
atgttaattt cactgtatgttgc ctgtatggggatc cggttccctt catcattttt catcttcacc 6300  
cccaaggactt gaatgatttcc agcaacttcc tccgggtgtatc cggccatgc caaaacttcg 6360  
tacaaaacacc acctttttatc tagggccatca gggccacggcc cccacccctcg atatattaa 6420  
agttccaggag agatgagggtt gcttccatgc accaggctgg ggtgtggacaca gcccgttgc 6480  
agtctgttcc tctgactatgtatcacttgc cggccacttcc cccagatcttcc agggcttgg 6540  
tgctgttcc cggggccgc atctggcccttgg gggactatgc ctgggtgtcc acacttgggg 6600  
cagccctgtcc tccacaccctt cccgttccatgc gcttccatgc ttccatccatc ttctttaacc 6660  
ctgggtgggc cgtgtttccatgc cgttactatgtatc tccatctgtcc cactgtgtatc tgctctccatgc 6720  
acgttagctcg cccgggtttccatgc cccatccatgc ggtgttccatgc tcccttccatc acacttccatc 6780  
cggttggaaat gggaggatgttgc tgcccttccatgc agacttgcgc ctctgtggatc gaaatctgg 6840  
cggtggccccc gatgtggatcc tccgggtccatgc ggtgtggatc tgacccatc tccatccatgc 6900  
tcccccgttcc ctgtatcttgc cccgggtggatccatgc cccgggtgtatc ttttttttttgc ttttttttttgc 6960  
tccacgttca gctgtgtatc tccatccatgc cccgggtggatccatgc tccatccatgc tccatccatgc 7020  
acggggggcggtt ggtggggccatgc ggcgttccatgc gggatccatgc cccggggccatgc 7080  
agtggccgttgc tccacacttgc tccacccatgc cccgggtggatccatgc tccatccatgc 7140  
ccggcccttc tccatccatgc cccgggtggatccatgc tccatccatgc tccatccatgc 7200  
ttccaccaacg actaaatccatgc tccatccatgc aagaccatgc atttccatgc tccatccatgc 7260  
gccccacatgc cccgggtggatccatgc cccgggtggatccatgc tccatccatgc tccatccatgc 7320  
ccgaccccccctt ctgtttttat ttaatagatca cccatccatgc tccatccatgc tccatccatgc 7380  
tttaacaaatc tgggtttttat ttaatagatca cccatccatgc tccatccatgc tccatccatgc 7440  
gaggaaatcatgc cccgggtggatccatgc tccatccatgc tccatccatgc tccatccatgc 7500  
tgcacccatccatgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc 7560  
agggggatgttgc ttagggggatccatgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc 7620  
cccttttacta aagccatgttgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc 7680  
cataggggatccatgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc 7740  
ccctggccatgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc 7800  
aaccctggccatgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc 7860  
atcccttccatgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc 7920  
aggagggttgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc tccatccatgc 7980





1 :  
 ctcttagtaat ctagtaatcc tttttttaaa ttgcctttag tactgcacaca ctgggcttct 3600  
 ttgttagt atttccctgc tgcgtctgtt ttgcctttt aatttatata tatatatata 3660  
 tttttttt ttttagaca gagtcttgc ctgcgcaca gggtagtgc agtgggtgtga 3720  
 5 tcacaggcgtca gtgtactt taccttctgg cctgagccgt cctcttacact cagccctctg 3780  
 agtagctgga actgcagaca cgcaccegtca caccctggta atttttaaat tttttctggta 3840  
 gacagggtct tgcgtgtgtg cccaggctgg tctcaaaactc ttggactcaa gggatccatc 3900  
 tacctcggct tcccaaaatgt ctgaattaca ggcatgagcc accatgtctg gcctaaattt 3960  
 caacactttt atatttttat agtgtggta tgcctgtta acagcatgtta ggtgaattttc 4020  
 caatccagtc tgacagtcgt tgcgtactt gataacccgtaa tttttttca tttttttgtc 4080  
 10 actagagacc cgcctgggtc actctgttcc tccacttgc tgcgtgtatgt cctctttttcc 4140  
 ttgtttctca ccaccccttgc ggttgcctatg tgcgtttctt gccgagttgtg tgcgtatctt 4200  
 ctcgttgcct cctggtaactt gggcatttgc tttttttctt ctttgccttag tgcgtacccccc 4260  
 tgatctttt attgctgtgt tttgtttttt ttttttttttttgc tgcgttctact ctgtcaccctt 4320  
 ggctggaggta taatggcaca atctcgctc actgcacactt ctgccttc ggttcaagca 4380  
 15 gttcttatttc ctcaaccccttgc tgatgtatgtt gggatttacagg cgcacccacac caccgcctggc 4440  
 taattttttt attttttagta gagataggtt ttcacccatgt tggccaggctt ggttcaaaac 4500  
 ttcgtacccctt aagtgtatgtt cccgccttgc cttccccacag tgcgtggattt acaggtgca 4560  
 gcccacccgtc cccggcatacc ttgtatctttt aaaaatgtt gtcggaaacattt gtcggccctt 4620  
 20 tccgtacccca taagacccctt aagtgtatttt agtgcacactt ctgccttc ggttcaagca 4680  
 tttttcccttgc tgacttagt ctatctcagg catcttgcata ccccccacaaatg ctaagcattt 4740  
 ttaatatttttgc tttccgtgtt ggttgcgttgc tgcgttttttgc cccccccttgc tttttcttcc 4800  
 tttgttcccttgc tgcgttgcctt tgcgttgcgttgc cccgccttgc ggttcccttgc ttttttttgc 4860  
 tgcgtgggttgc ttgtgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 4920  
 atggcatctca ggcacgttgc gggaccccttgc cttatgtatgc acatgttgc acatgttgc 4980  
 25 tcacagggttttttgc ggcgggttcaatc ttttttttttttgc ttttttttttttgc ttttttttttttgc 5040  
 ctttagccatgttgc ggttgcgttgc aacgttgcgttgc cccgccttgc ttttttttttttgc 5100  
 catgttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 5160  
 ggcacccgttgc tgacgggttgc ggcgttgcgttgc ggttgcgttgc ttttttttttttgc 5220  
 30 aaccaggaca aaggatgagg ctccgagccg ttgtcgccca acaggagcat gacgttgc 5280  
 atgtggatata ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 5340  
 ctttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 5400  
 atgtatccaaatccatctgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 5460  
 ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 5520  
 35 gtttgcgttgc ggcacatttgc caccacttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 5580  
 ctcaaaaaaaa aaaaaaaaaaaa aaaaaaaaaaaa aatttttttttttgc ttttttttttttgc ttttttttttttgc 5640  
 aagaaaaatgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 5700  
 atcattttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 5760  
 ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 5820  
 ggcctccatgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 5880  
 40 ccccttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 5940  
 ccccttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6000  
 gcccttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6060  
 gcccttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6120  
 ggccttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6180  
 45 ggccttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6240  
 agggcccttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6300  
 cagacgggttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6360  
 ggttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6420  
 ggttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6480  
 50 ggggttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6540  
 gtttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6600  
 gtttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6660  
 ggttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6720  
 ggttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6780  
 55 aggttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6840  
 caggttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6900  
 caggttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 6960  
 gcaagggttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 7020  
 gcaagggttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 7080  
 60 ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 7140  
 ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 7200  
 ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 7260  
 ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 7320  
 gtttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 7380  
 gtttgcgttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc ttttttttttttgc 7440



```

<210> 7
<211> 687
<212> DNA
<213> Homo sapiens
5
<400> 7
gtggctgtgc tttggtttaa cttccctttt aaacagaagt gcgttgagc cccacatgg 60
gtatcagctt agatgaaggg cccggaggag gggccacggg acacagccag ggccatggca 120
cgcgccaaac ccatttgcgc gcacagttag gttggccgagg tgccgggtgcc tccagaaaaag 180
10 cagcgtgggg gtgttaggggg agtccttggg gcagggacag gctctgagga ccacaagaag 240
cagccggggc agggcccttggg tgacgcacgg cccggaggcc tggatccgtg tccctgtgt 300
gtgcgcagcc tccgtgcgtc tccgcttaa gggcccccggg accaggccac gactgcccagg 360
agcccacccgg gctctgagga tccctggaccc tggcccccacgg ctccctgcacc ccacccctgt 420
ggctgcgggtg gctgcgggtg ccccgctatc tgaggagagt gtgggggtgag gtggacagag 480
gtgtggccatg aggatccctgt gtgcacacaca catgcggccca ggaacccgtt tcaaaacaggg 540
15 tctgaggaag ctgggggggg ttcttaggtcc cgggtcttggg tggctgggggca cactggggg 600
gggtgtcttc tccccctgggt ccctatgggt gggtggggcac tggccggat ccacccctct 660
gactgtctcc catgcgttcc ccgcacag 687
20
<210> 8
<211> 494
<212> DNA
<213> Homo sapiens
25
<400> 8
gtgggtgccc gggacccccc tgagcagcccc tgcgtggaccc tggggagtggc tgcctgtattg 60
gcacccatgtt tgggtggag gaggacttcc tgggtggggcc gcaggaggtg cagggtaccc 120
tgcactgtt gaggacacac ctggcaccta gggtggaggc ctccagccctt tccctgcagca 180
catggggccg actgtgcacc ctgactgccc gggctccat tcccaaggag ggtcccactg 240
30 gattccatgtt tccgtcagag aaggaaacccgc aacggctcag ccacccaggcc cccgtgcctt 300
gcaccccccctt cctgagccag ggggtctctg tccctgagget cagagggggg acacagcccg 360
ccttgcctt ggggtcttgg gttgggtgggg tcaagagagag agtgggggac accggccaggc 420
caggcccttga gggcagagggt gatgttggag tttctgcgtg gcaactgtca gtcttctcgc 480
ctccactcactc acag 494
35
<210> 9
<211> 865
<212> DNA
<213> Homo sapiens
40
<400> 9
gtaagggttca cgtgtgatag tgcgtgtcccg gatgtgtgtc tctgggatat gaatgtgtct 60
agaatgcagg cgtgtctgtg atgcgttctt gtgggtggagg tacttccatg atttacacat 120
45 ctgtgtatgt cgtgtgtggc acgtgtgtgt cgtgtgtcat gtatctgtgg cgtgcattt 180
tgcgtgtgtgt gtgtgtgtgg cacgtgtgtg tccatgggtgt gtgtgtctgt ggtgtgcattg 240
tgcgtgtgtc tgcgtacacgt gcatgttccat gctgtgtgtc gcatgtctgt gatgtgcctt 300
tttgcgtgtgt gtgtgtgtcat gtgtccgttca gatgtgtgt tctatggcat ggggtgtgt 360
ggcccttggg ctttactctt tccctccatca ggcgtgtcc gcacattgtt cctcactgtc 420
tcgggtgtgt gtttggggag ctccacatcc aggggttccatc ctcttagcat ggggtccccct 480
50 gtcctgtcactt aggggtgtggc cttggagact gtaaggccagg tttgagagga gagtagggat 540
gtcggtgttca ctttcccttgg gcccctggcc ccccaaggaccc ccagtttgcgc ctatggccgc 600
tccatggat ataggaaaggc tgattccaggc ctgcgtcccc gggacacact cctcccaagag 660
cggccgggggg ctttgggggtt cggcagggggtt gaaaggggcc ctgggtttgg gttttccaccc 720
55 agtgggtcatg agcacgttgg aggggttcaagc cttcaaaatc gtgcaggcc ggggtgtcaga 780
ggtaagaag tttcccttgg gtttgcgtt ggggagaggc acatgtggaa acccacaagg 840
acctttttt ctgactttttt gagct 865
60
<210> 10
<211> 3782
<212> DNA
<213> Homo sapiens
65
<400> 10
tgcgtgtgtt gttttcatgt gtgggtatgg tggggatctg tgggtttgggt tttttatgtgt 60
gggttacacac agatgttcaag gtcgttttcc tttctgtatgtt ggggtttgtccatg gtcgttcaac 120

```

agctttttatg aggagaccat atctccctt gaactatggt cgggttata gtaagtcagg 180  
 ggtgtggagg cttcccttgg gtccttctt ctgttttttc cactctgggg tcgtgtggtg 240  
 cctgtgtgg tggtgcccg gttggcaggg ctccagggc tccttgggtt cattggctg 300  
 5 gatgtggccc tggtaacgt cctgttctgg aattccctgg cgaatggag gcttttttc 360  
 tttttttttt tttttttttt tgataacaga gtctctctt tttttggcca 420  
 ggctggagtg gtttggctg atcttggctc actgcaacat gtcttccctg agtcaagca 480  
 atcttcttgc ctcaagctcc caagtagctg gaattatagg cggccaccac catgtact 540  
 aatttttta attttagtag agacgaggtt tctccatgtt ggccaggctg gtctcaact 600  
 cctgacccca ggtgatccctt ccacccctggc ctcccaaagt gctggatgaa cagggtgtgaa 660  
 10 ccggccggcc cggccggagac tcgttctctg cagatccctg gagatctgca gcgatagctg 720  
 cctgcagccct tggtgctgac aacccctggc ttcccttccctt agtctctgtt aggggttctt 780  
 ccatttcatg actcttttca cagaaggtt tcaatgttgc tgattttccctg gctgtttctt 840  
 gctgttggt tggatgttgc ttatcgatgg cctccatcca tttccctttag gctttgttta 900  
 ttgttggttt tccggctctt tgaaggaaaa gtttggatgg tggatgtttt aactttttt 960  
 15 tctaaacaag catctgaatg tggcgatttt cctcttaaage agggatccctgg aggccccctgg 1020  
 ctgtggatgtt gCACCGTCTT gggggctttt aggaaccccg cgcacagccg gaggcttaggt 1080  
 ggggtgtggg gacccagctg tcccgccctg gcccccccccc tctcagatca gcaatggcat 1140  
 gccgtgtctca gagggccaca caccctactg agaactgtgc gtgagggggg tcttagattct 1200  
 20 gtgtcttta tggaaatcta atgcctgtatg atctgagggtt gaaaccgtttt ctccaaac 1260  
 catcccccttc cccactgtctg tcctgtggaa aaatctgtt ccacggaaacc agtccctgtt 1320  
 accacaatgg ttggggaccc tggatcttcaag acctgttca gcaatgttctc gtcagtgttt 1380  
 atatattggc ttttctgtt tggatccatgg aatattacgg atttctgttga tggatccctc 1440  
 cgacccatcaga cccatggcc attttggggc gtgttgcctg ctccctgggtt gggaaagggtt 1500  
 caggccccat gtacccctt gttactgcctt tccagggttgg ttcctcagggtt tgaatcgat 1560  
 25 tcgtatgtgtt ttttagccac gggccctggcc ccagcttctg ggggtgggg aacatgttca 1620  
 agcacacatgg caccgtgcgc tccttttgcgtt gcttcaacaag ctcgaggccct cctgtgtccg 1680  
 tggatgtgtt tggatcgatgtc ctgtatccatgtt cctgttgcgtt ggacccgggg gcttagcagg 1740  
 tcccgtatgtt aatgtcaagc gtccctgggg agtctcgatgg ataggaggtt ggggtgtccgg 1800  
 30 ttcctctcc gctgttttcaatcttcc tcctgttgcgtt gtggatgcac ctgcattccct 1860  
 gcaatccctc cagcaactggg ctggagggc cccggagctc gatgtgcact tggccacat 1920  
 gactgtggat ggcagtcgtt cacccggggc tggatgtgtt tggatgttgg tggcggttgg 1980  
 tcacagggtt ctgatgtgtt tggactgttgc atggccgtcg tggggcttga tggatgtact 2040  
 gtggatgtccg tggatgtgggtt ctgatgtgttgc tggactgttgc atggccgtcg tggggcttga 2100  
 tggatgtact gtggatgtccg tggatgtgggtt ctgatgtgttgc tggatgttgg ggcgggtcg 2160  
 35 gggatgtatgtt tggatgtactgtt ggtatggcgtt cgtggggctt gatgtgttgg tggatgttggat 2220  
 ggcggatgtatgtt tggatgtactgtt ggtatggcgtt cgtggggctt gatgtgttgg tggatgttggat 2280  
 gactgtggat ggcggatgtatgtt tggatgtactgtt ggtatggcgtt gatgtgttgg tggatgttggat 2340  
 ctgatgtgtt tggatgttgcgtt atggccgtcg tggggcttga tggatgttgg tggatgttggat 2400  
 cggtcgatggg tggatgttgcgtt atggccgtcg tggggcttga tggatgttgg tggatgttggat 2460  
 40 ctgtggatgg tggatcgatca caggggtctg atgtgtgggtt actgtggatg ggcggatgtccg 2520  
 ggtctgtatgtt tggatgtactgtt tggatgttgg tggatgttgcgtt ggtatgttggact 2580  
 gtggatggcc tggatgtgggtt ctgatgtgttgc tggactgttgc atggccgtcg tggccggggg 2640  
 tggatgtgtt tggatgttgcgtt atggccgtcg tggggcttga tggatgttgg tggatgttggat 2700  
 ggatggccgtt cgtggggctt gatgtgttgg tggatgttgg ggcggatgtccg gggatgtatgt 2760  
 45 tggatgtact gtggatgtccg tggatgtgggtt ctgatgtgttgc tggatgttgg ggcggatgtccg 2820  
 gggatgtatgtt tggatgtactgtt ggtatggcgtt cgtggggctt gatgtgttgg tggatgttggat 2880  
 ggcggatgtatgtt tggatgtactgtt ggtatggcgtt cgtggggctt gatgtgttgg tggatgttggat 2940  
 gtggatgtact tggatgttgcgtt tggatgttgg tggatgttgg tggatgttgg tggatgttggat 3000  
 gggatgtatgtt tggatgttgcgtt tggatgttgg tggatgttgg tggatgttgg tggatgttggat 3060  
 50 gatggccgtt tgggggtctg atgtgtgggtt actgtggatg gggatgttgg tggatgttggat 3120  
 ggtgactgtt gatggccgtt tgggggtctg atgtgtgggtt actgtggatg gggatgttgg tggatgttggat 3180  
 acagggttctt gatgtgttgg tggatgttgg gggatgttgg tggatgttgg tggatgttggat 3240  
 gtggatgtccg tggatgtgggtt ctgatgtgttgc tggatgttgg tggatgttgg tggatgttggat 3300  
 tggatgtact gtggatgtccg tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttggat 3360  
 55 gtccatgggg tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttggat 3420  
 actgtggatg gggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttggat 3480  
 tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttggat 3540  
 tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttggat 3600  
 acgttgcgtt ctggccccc cggcccccgtt tcccaaaaca gaaatcttcc agggccgttcc 3660  
 tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttggat 3720  
 60 agtggccatc tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttggat 3780  
 agtggccatc tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttgg tggatgttggat 3782





5                    gggattgtcc aatgtggtcc ccctcaaggg cgccccacag agccgggtggg cttgttttaa 660  
 agtgcgatc gacgaggggac gagaaaccc taaagctgtt aaggaaaccc tcagaaaatg 720  
 tggccgcacag ggggtggttt aggtgtttt ctgggtgtg tttgtaaaaa cccatggaa 780  
 cccgcctcc aagtccaccc tccagggtca ccctccaggg ccgcctggg ctgggggtat 840  
 gcttggcgtt ctttgtcccg cagcccgag cacagcagc tgcacatt taaaatccact 900  
 aagatttact cggggggagc ccagggtcca agcaactgag ggctcaggag tcctgaggct 960  
 gctgagggga cagagcagac ggggaacgtt gtttctgtt ggcaagtcc tgaggggtct 1020  
 ggcgcagggg gtggctcaga gtgtatgtt gggtccccc gggggcagaa ctctgttct 1080  
 gatgagtcgg cagccatgtt acaggaaagg gtggccacag ggagctgggaa atgcaccagg 1140  
 10                ggagctgcgc agtggccga ggtcccgcc ccaggccaca gggaggcag ggggacgccc 1200  
 gggccacacag cagagccgc aggaaggaa ggggtatggcc aggcacagc agaggcttac 1260  
 gggcacaggg gggctccctg agctgggtga gcgaggctca tgactcggcg agggacaccc 1320  
 cttgacgtga agctgacgac tgggtgtcc cagcttacag cccagccagg tcccgccct 1380  
 gaggcaggaa tcagaaccc ccccttgc taaaggccacag catgtccctt caggccatcc 1440  
 15                aggagaaaaac aggccaaatgg gttgagaaac gtctttaaaag aagggtggat ggtggcaatt 1500  
 tcttttccat attttatgtt gccccggacc acagatgtt ctataacggg atttgtgtt 1560  
 tgccatgggg acacatgaga tggccatca cagaggccac tgggtgtca cctccatct 1620  
 gatgtctggc tggccctggg ccaggccagg ttcttgcgtt ctcacccatcc tgcctctgccc 1680  
 20                gggagacagg gaaaggacccc cgaagtttgg agcagggtgtt gggtccaggct cctcagatg 1740  
 cctggccaggc ccagccaccc gctccaaatc accatcttc tgggttttc caaagcattt 1800  
 aacaagggtt tcaagggttacc tccctgggtga cggcccccga tccctgggtt gacattggccc 1860  
 ctctgccttaa g 1871

25                <210> 15  
 <211> 3801  
 <212> DNA  
 <213> Homo sapiens

30                <400> 15  
 gtgagcgcac ctggccggaa gtggaggttgc tgcccccgtt gggcagggtgc tgctgcagg 60  
 ccgttgcgtc caccctgttcc tccgtgtggg gcaggcgact gccaatccca aagggtcaga 120  
 ggcacagggg tgccctctgtt cccatctggg gtcgagcaga aatgcattt tctgtgggg 180  
 tgagggtgtt cacaacggga gcagttttt gtgtatttt ggtaaaaggaa aatgggtgcac 240  
 cagaccttggg tgcactgagg tgcacttgc aagcgttgc gatccgaaacc caagacgccc 300  
 35                gggccctgtt gggcgtgtt ctcttcaacc cgaacacagg ggcctgtgtt ggcgttgc 360  
 cctctgaacc ctagacccctg gggccctgtt gggcgtgtt ctcttcaacc ccagagactt 420  
 caggccctt ttgggcgttga gtcttccctg tgcgtggccca acatccaaag gtcctatcc 480  
 agtctacagg atgcattgtt ttcattatgtt cgtgtgttcc accatggggcaggccatgg 540  
 tgggggggg gtcacttacaa aattctgggg tctttttcc ccaagccccg agagctcaag 600  
 40                gccccgttcc aggcttccatc accaaatgtt tgcgtgttcc cacaatgttca gaatctgtt 660  
 ctgttttccatc tatgataaa aagtatccatc attccaggaa gggcaagggtt gtcacacc 720  
 ataatcccaatc cactttggg gggccagggtt ggtgttccatc ttggggccag ggttttgggg 780  
 ccaacccatc caacatgtt accatccatc tctacttaaa aaatataaaa attagctgg 840  
 cctggtggca caccgttgc tgcggccgtt tgaggccaggaa gaatcatttgc 900  
 45                aacccaggag gcagagggtt cgtgtgttcc agatcacacc actgcactcc agccctggca 960  
 acagactgtt accttcatcc taaaatccatc taaaatccatc caccatccaa accatccatc 1020  
 ggacagggtt ttttttccatc tgcgtgttccatc ttttttccatc tgggtgttcc ttagaggcc 1080  
 gaactggggg tgccttccctg tggaaaggccac accttcatcc tgggtgttccatc ttttttccatc 1140  
 50                tgggtgttccatc accatccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 1200  
 tggacttttc ttttttccatc aatgtttccatc ttttttccatc ttttttccatc ttttttccatc 1260  
 tggacaccctt cgtgtgttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 1320  
 ggtgtgttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 1380  
 ggatccctgttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 1440  
 ctggccctccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 1500  
 55                cgccttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 1560  
 ctttttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 1620  
 gcacacgtt accatccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 1680  
 gtttttccatc aacacatccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 1740  
 catcgttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 1800  
 60                gtgtgttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 1860  
 ggggtgttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 1920  
 gaagaaaaaca ggcacaaatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 1980  
 ttcttgccttccatc gatgttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 2040  
 gatggacaga acaatagaaac ttttttccatc ttttttccatc ttttttccatc ttttttccatc 2100  
 65                gttatgtggca cagctgttccatc ttttttccatc ttttttccatc ttttttccatc ttttttccatc 2160







5 taggttagact gtcaaatttc agagcaaaat gaaaataaca aagttttaaa gggaaagaaa 2340  
aaccggact ctttggactt ccttaggcct gaacttcatc tcaagcaget tccttcaca 2400  
gacaaggctg tatggagcga gtgagttcaa agcagaaagg gaggagaagc aggcaagggt 2460  
ggaggctgtg ggtgacacca gccaggaccc ctgaaaggga gtggtttt tcctgcctca 2520  
gccccacgct cctgcccgtc ctgcacccgtc tgtaaccgtc gatgtgggtg ccaggtgccc 2580  
acctggaaag gatgtgtgc agggggcttg ccaaactttg gtgggttca gaagccccag 2640  
gcacttgtgg caggcacaat tacagccctt ccccaaagat gcccacgtcc ttcttcgtga 2700  
acctgtgaat gtgtcacccg caaggcagag gctggtgaag gctgcagggtg gaatcacggc 2760  
tgccagtcag ccgatcttaa ggtcatctg gattatctgg tgggcctgat atggccaca 2820  
gggtcccttag aagttagaga gggaggcagg ggagagtcag agaggggacg tgagaaggac 2880  
caactggccac tgctggctt gagatggagg agggggtccc cagccaaaggga atggggcag 2940  
ccgctccatg ctggaaaagc aagcaatctt ccccggtctt gagggcacac ggcctgccc 3000  
acgcctcgat ttcaggccag tgggacctgt ttcagcttcc cgccctccag agctgtaa 3060  
tgatgcgtt gtgttcagcc actaagctgc agtattcgat cacagcagca aatggaaatag 3120  
10 cagtagcaggg aaatgaatac agggacagtt ctcagagtga ctctcagccc acccctggg 3179  
15